



**Système complet d'acquisition vidéo, de suivi de
trajectoires et de modélisation comportementale pour
des environnements 3D naturellement encombrés :
application à la surveillance apicole**

Guillaume Chiron

► **To cite this version:**

Guillaume Chiron. Système complet d'acquisition vidéo, de suivi de trajectoires et de modélisation comportementale pour des environnements 3D naturellement encombrés : application à la surveillance apicole. Traitement du signal et de l'image [eess.SP]. Université de La Rochelle, 2014. Français. NNT : 2014LAROS030 . tel-01264108

HAL Id: tel-01264108

<https://theses.hal.science/tel-01264108>

Submitted on 28 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ DE LA ROCHELLE

ÉCOLE DOCTORALE S2IM

Laboratoire L3i (Informatique, Image et Interaction)

THÈSE présentée par :

Guillaume CHIRON

soutenue le : **28 novembre 2014**

pour obtenir le grade de : **Docteur de l'Université de La Rochelle**

Discipline : **Automatique, Image et Signal**

**Système complet d'acquisition vidéo, de suivi de trajectoires
et de modélisation comportementale pour des environnements
3D naturellement encombrés**

Application à la surveillance apicole

JURY :

Didier DEMIGNY

Jean-Philippe DOMENGER

Yves DEMAZEAU

Alain TREMEAU

Pierrick AUPINEL

Michel MÉNARD

Petra GOMEZ-KRÄMER

Professeur, Université de Rennes 1, Rapporteur

Professeur, Université Bordeaux 1, Rapporteur

Directeur de Recherche, CNRS, Examineur

Professeur, Université de Saint-Étienne, Examineur

Directeur d'unité, INRA Magneraud, Invité

Professeur, Université de La Rochelle, Directeur de thèse

Maître de conférences, Université de La Rochelle, Coencadrante de thèse

Remerciements

Tout d’abord, je tiens à exprimer mon immense reconnaissance à mes encadrants, **Michel MÉNARD** (Professeur à l’université de La Rochelle), et **Petra GOMEZ-KRÄMER** (Maître de conférences à l’université de La Rochelle) que je ne saurais trop remercier. Durant ces 3 années, j’ai vraiment apprécié votre accompagnement que je qualifierais d’exceptionnel, tant sur le plan humain que sur le plan scientifique. Au fil de nos nombreux échanges, vous m’avez transmis cette véritable passion pour la recherche et l’enseignement, et cela, dans une ambiance détendue de travail mais non moins rigoureuse. Selon moi, cela est la recette idéale pour mener une thèse dans les meilleures conditions. J’espère que nous aurons, à l’avenir, le plaisir de collaborer à nouveau ensemble.

Je remercie les membres du jury d’avoir accepté d’évaluer ces travaux de thèse. À savoir :

- **Didier DEMIGNY** (Professeur à l’université de Rennes 1) et **Jean-Philippe DOMENGER** (Professeur à l’université Bordeaux 1) pour leur travail en tant que rapporteurs ;
- **Alain TREMEAU** (Professeur à l’université de Saint-Étienne) qui fut le président du jury et **Yves DEMAZEAU** (Directeur de Recherche CNRS) pour leur rôle d’examineurs ;
- **Pierrick AUPINEL** (Directeur d’unité, INRA Magneraud) en tant qu’invité pour son regard d’expert sur l’application apicole ;
- mes encadrants, **Michel MÉNARD** et **Petra GOMEZ-KRÄMER** qui ont également pris part à ce jury.

Je voudrais également remercier toute l’équipe du département informatique de l’IUT de la Rochelle, en particulier **Annick LASSUS**, **Cyril FAUCHER**, **Laurent BERNDT**, **Marie-Hélène VERRONS** et **Philippe COULAUD** avec qui j’ai eu l’occasion de partager une partie de mon service de DCACE et d’ATER. Avec vous, j’ai découvert le métier d’enseignant. Vous resterez pour moi des modèles.

Quelques mots aussi pour remercier **Jean-Marc OGIER** (Directeur du laboratoire), et l’ensemble des chercheurs rattachés au L3i pour leur accueil chaleureux, mais surtout pour avoir bâti au fil des années cet environnement de travail agréable qui permet à nous, ex-doctorants et futurs docteurs, de travailler dans les meilleures conditions. J’en profite pour remercier les collègues volontaires qui ont contribué à l’élaboration de bases d’images annotées pour le projet Abeille.

Un grand merci aux personnes travaillant au sein de l’axe environnement (monté au L3i en 2011). Je pense notamment à **Bruno LESCALIER** et **Olivier CHAPPE** avec qui nous avons réalisé un travail considérable d’ingénierie et de développement sur ce projet Abeille.

Mes remerciements vont à présent aux biologistes de l’INRA Magneraud, pour avoir mis à notre disposition des ruchers expérimentaux, et surtout pour la confiance qu’ils nous ont accordée lors des montages des projets interdisciplinaires EPERAS, RISQAPI et FREDD. Merci à **Vincent BRETAGNOLLE** (Directeur de Recherche CNRS, CEBC) pour avoir ouvert la voie de ces collaborations qui ont permis d’offrir un cadre applicatif original et stimulant à cette thèse. Merci en particulier à **Fabrice REQUIER** (Docteur en Ecologie

de l'abeille) pour ses conseils et son accompagnement sur le terrain lors des manipulations biologiques chronophages qui se sont déroulées durant l'été 2013.

En termes d'accompagnement sur le terrain, je tiens également à remercier **Quentin ROME** et **Judith ANDRIVOT** du Museum National d'Histoire Naturelle pour nous avoir fait bénéficier de leur expertise lors de la campagne d'acquisition vidéo de l'automne 2013 qui s'est déroulée dans les ruchers du Bas Mairé (à Aiffres).

Un merci également à **Benjamin POIROT** (Co-Fondateur APILAB) pour nous avoir accompagné et conseillé au démarrage cette thèse, et également pour nous avoir mis à disposition son matériel et ses ruches.

Je remercie aussi **Yves LE CONTE** et **Didier CRAUSER** (de l'INRA Avignon) d'avoir accepté notre invitation et de s'être déplacé à la Rochelle pour découvrir nos travaux autour de l'abeille.

Aristote estime que l'homme est un animal social. Je pense que tout comme l'abeille, nous avons besoin de notre colonie pour subsister. Ainsi, je tiens à remercier ma colonie, et notamment :

- ceux qui me côtoient au quotidien et qui me donnent la force d'avancer : **Axel, Antoine, Bruno, Bich, Christophe, Clément, Cyril, Dounia, Élodie, Joseph, Marcela, Maroua, Mickaël, Olivier, Omar, Romain** (merci pour ta relecture du manuscrit), **Rouaa, Sébastien, Sophea, Vincent** ;
- mes amis proches mais géographiquement éloignés qui ont tout de même été à mon écoute et m'ont aidé à me changer les idées le temps d'un weekend ou d'une simple soirée ;
- ma famille qui m'a soutenu dans mes choix et m'a aidé à prendre du recul dans les moments difficiles ;
- à ceux qui se reconnaîtront que je n'ai pas mentionné.

Je conclus ces remerciements par une tendre pensée envers toi, **Qian**, qui partage mon quotidien depuis déjà plusieurs années. Ton soutien a été sans faille, même pendant ces derniers mois difficiles "tête dans le guidon". Tu as su m'offrir ces bouffées d'oxygène et d'exotisme qui m'ont permis de m'évader de cette bulle si singulière qu'est la thèse.

Financements

- Contrat doctoral financé par le Conseil général de la Charente-Maritime (17) ;
- Matériel financé Fonds européen de développement régional (contrat : 35053) ;
- Projets interdisciplinaires :
 - RISQAPI : FranceAgrimer (contrat 14-03R), labellisé ANR ;
 - EPERAS : Région Poitou-Charentes.

Table des matières

Remerciements	i
Introduction générale	5
Chapitre 1. Acquisition stéréovision et détection de cibles	15
1.1 Introduction	16
1.2 Dispositif d'acquisition	16
1.2.1 Contraintes applicatives	17
1.2.2 Choix du système d'acquisition 3D	18
1.2.3 Configuration du dispositif stéréovision	24
1.3 État de l'art sur la détection de cibles	25
1.3.1 Détection par modèle d'apparence	25
1.3.2 Détection par le mouvement	26
1.3.3 Détection sur des données RGB-D	27
1.3.4 Détecteurs utilisés pour les insectes	27
1.4 Solution proposée : une segmentation hybride	28
1.4.1 Complémentarité des données d'intensité et de profondeur	28
1.4.2 Algorithme HIDS (Hybrid Intensity Depth Segmentation)	31
1.4.3 Extraction des cibles	37
1.5 Évaluation de la méthode de segmentation HIDS	38
1.5.1 Élaboration de la vérité terrain	39
1.5.2 Résultats	41
1.5.3 Synthèse sur l'influence des paramètres	43
1.6 Conclusion	44
Chapitre 2. Suivi 3D multi-cibles	45
2.1 Introduction	46
2.2 État de l'art des approches probabilistes	47
2.2.1 Contexte théorique	47
2.2.2 Formalisation du problème	48
2.2.3 Méthodes classiques de filtrage	49
2.2.4 Variantes existantes, pour quels contextes ?	52
2.2.5 Association de cibles pour le suivi par filtrage de Kalman	56
2.3 Solutions théoriques proposées	58
2.3.1 Filtrage de Kalman 3D et mesures 2D/3D	59
2.3.2 Suivi contraint par un modèle de scène	62

2.3.3	Suivi contraint par une base de données de trajectoires	65
2.4	Préparation de la mise en œuvre pour le suivi d'abeilles	68
2.4.1	Présentation de l'application	68
2.4.2	Configuration du filtre de Kalman	71
2.4.3	Modélisation de la planche d'envol	71
2.4.4	Construction d'une base de données de trajectoires d'abeilles en vol . .	73
2.5	Méthode d'évaluation et résultats	74
2.5.1	Mise en place d'une vérité terrain	74
2.5.2	Définition des modèles de scénario	75
2.5.3	Métrique pour l'évaluation	75
2.5.4	Résultats sur différents modèles de scénario	76
2.6	Conclusion	82
Chapitre 3.	Modélisation comportementale	85
3.1	Introduction	86
3.2	État de l'art sur les modèles Bayésiennes non-paramétriques	87
3.2.1	Positionnement sur la modélisation comportementale	87
3.2.2	Présentation succincte des modèles Bayésiens non-paramétriques	88
3.2.3	Extensions du Processus de Dirichlet	92
3.3	Modèle MLC-HDP pour des trajectoires issues de séquences vidéo	97
3.3.1	Structure de données	97
3.3.2	Clustering selon trois niveaux	98
3.3.3	Signatures de trajectoires	100
3.3.4	Critère d'arrêt	102
3.4	Simulation Multi-Agents pour la validation de l'approche	103
3.4.1	Pertinence du couple SMA / MLC-HDP	103
3.4.2	Génération de la vérité terrain	103
3.4.3	Étude préliminaire	105
3.4.4	Étude approfondie	110
3.5	Conclusion	118
Chapitre 4.	Applications à la surveillance apicole	119
4.1	Introduction	120
4.2	Dispositifs existants pour la surveillance de ruches	121
4.2.1	Compteurs d'entrées/sorties	121
4.2.2	Analyse des trajectoires	123
4.2.3	Autres approches	124
4.3	Protocole expérimental	124
4.3.1	Contraintes de terrain	125
4.3.2	Calendrier des interventions	126
4.3.3	Dispositif d'acquisition	127
4.3.4	Préparation des données et prétraitements	128
4.4	Traitement des séquences vidéo	130
4.4.1	Détection des abeilles	130
4.4.2	Calcul des trajectoires	132
4.4.3	Extraction des caractéristiques	132

4.4.4	Modélisation MLC-HDP pour RISQAPI	135
4.4.5	Synthèse sur les performances	139
4.5	Identification des trajectoires d'abeilles	141
4.5.1	Mise en correspondance des histoires de vie avec les trajectoires	141
4.5.2	Détection visuelle des puces RFID	142
4.5.3	Nettoyage des données RFID	143
4.5.4	Modalités d'application pour EPERAS	145
4.6	Travaux annexes sur la détection du frelon asiatique	146
4.6.1	Détection du frelon par sa taille	146
4.6.2	Analyse de l'activité sur la planche d'envol	147
4.7	Conclusion	148
Conclusion générale et perspectives		151
Annexe A. Outils d'annotation		157
A.1	Annotation de cibles	157
A.2	Annotation de trajectoires	158
A.3	Détournement de la planche d'envol	159
Annexe B. SMA de "Colonie d'abeilles"		161
Annexe C. Fiche d'annotations		163
Annexe D. Détection visuelle des marqueurs		165
Publications		167
Couverture médiatique		169
Liste des Abréviations		171
Table des Figures		172
Liste des Tableaux		176
Liste des Algorithmes		177
Bibliographie		179

Introduction générale

Contexte de la thèse et motivations

Le développement durable et la sauvegarde de notre environnement sont au cœur des enjeux gouvernementaux et de l'opinion publique. La réglementation en la matière se durcit d'années en années. En particulier, les sites à risque technologique comme les incinérateurs ou les zones de stockage des déchets doivent respecter des normes strictes et sont soumis à des contrôles réguliers de l'état sanitaire de leur environnement.

Initialement, ce sujet de thèse fut motivé par le développement d'outils permettant de s'appuyer sur l'observation des abeilles pour la surveillance environnementale des sites à risque. On parle alors d'abeilles sentinelles de l'environnement, c'est-à-dire que la présence de celles-ci et leurs activités dans le milieu reflètent la variation des conditions environnementales de ce dernier. Une intoxication d'abeilles se traduit soit par une mortalité directe de celles-ci, soit par une modification de leur comportement si la quantité de polluant n'atteint pas une dose létale [Henry 2012]. Le laboratoire L3i de la Rochelle, en partenariat l'INRA Avignon ainsi qu'avec des entreprises locales et nationales (ApiLab, CANSIE et AZIMUT), travaillent depuis quelques années sur plusieurs projets¹ allant sur ces problématiques.

De fil en aiguille, alors que nous exposons nos travaux préliminaires lors du colloque FREDD² du 28 juin 2012, une rencontre avec le directeur du Centre d'Etudes Biologique de Chizé (CEBC), Vincent Bretagnolle, nous inspira de nouveaux objectifs : étudier les effets toxicologiques de l'exposition, à des doses sublétales, de produits phytosanitaires sur l'abeille domestique. Cette idée fut rapidement concrétisée sous la forme du projet EPERAS³ en collaboration avec l'INRA Magneraud et le CEBC. Puis, par la suite, une seconde collaboration fut montée avec le Muséum national d'Histoire naturelle National dans le cadre du projet RISQAPI⁴. Ce dernier a permis d'élargir le champ applicatif de cette thèse à nouveau sur des problématiques bien réelles. Il s'agit cette fois de mesurer l'impact sur les abeilles de la prédation par le frelon asiatique.

Le déclin actuel de l'abeille domestique (*Apis mellifera*) n'est plus à démontrer. La littérature y fait référence sous les termes de "syndrome d'effondrement des colonies d'abeilles" (ou CCD pour Colony Collapse Disorder). Ce phénomène fait l'objet de nombreux travaux de recherche, autant publics que privés, qui sont au centre d'énormes enjeux économiques, en raison notamment de l'indispensable service de pollinisation rendu par ces insectes. Plusieurs

1. PRIM'INNOV (développement d'un compteur d'abeilles) ; PRESERVE (développement d'une plateforme de surveillance environnementale) ; projet de compteur embarqué avec la société AZIMUT.

2. Fédération de Recherche en Environnement pour le Développement Durable créée dans le cadre du PRES Limousin Poitou-Charentes.

3. Projet EPERAS financé par la région Poitou-Charentes dans le cadre de l'appel à projet "Recherche et santé environnementale, 2012/2015"

4. Projet RISQAPI (labellisé ANR) : lien spatial entre usage des pesticides et risques d'affaiblissement des colonies d'abeilles en système de grandes cultures. Porteur du projet : INRA Avignon.

facteurs agissant en synergie sont pointés du doigt par les biologistes. Nous ne citons ici que les principaux :

- une prédation grandissante du frelon asiatique [Arca 2014] ;
- une exposition toujours plus importante à des produits phytosanitaires variés soupçonnés de provoquer des effets cocktails, notamment les néonicotinoïdes [Vidau 2011] ;
- la dispersion dans le milieu du "Varroa destructor", un acarien s'attaquant aux abeilles adultes, mais également aux larves et aux nymphes [Sumpter 2004].

De récentes études ont montré l'impact prépondérant des pesticides parmi la synergie factorielle de pression agissant sur le déclin de l'abeille domestique. Dans [Henry 2012], les auteurs ont mis en évidence que les pesticides troublent le comportement de vol des butineuses par désorientation, diminuant considérablement la probabilité de retour à la ruche. Ces résultats ont pu être obtenus grâce à l'utilisation d'un nouvel outil ingénieux permettant de retracer les entrées et sorties des abeilles au niveau de la ruche, à savoir le dispositif RFID (Radio-Frequency Identification) [Streit 2003]. Cependant cet outil ne retranscrit ni le comportement de vol des butineuses, ni le décodage des danses des abeilles [Von Frisch 1965]. Le seul outil capable de fournir ces informations est le radar harmonique [Osborne 1999]. Mais ce dernier outil biaise le comportement de vol de l'abeille par l'ajout d'un traceur volumineux sur l'abeille et ne permet de suivre que quelques individus à la fois. Les biologistes sont donc encore à l'heure actuelle démunis d'outils permettant l'étude du comportement de vol des 1000 à 10000 butineuses présentes dans la colonie.

Il existe dans la littérature informatique des approches décrivant le suivi biologique de ruches par vidéosurveillance. Par exemple, [Blois 2011] décrit une méthode de comptage des entrées/sorties d'abeilles à l'aide d'une caméra placée à l'entrée de la ruche. Ce travail s'inspire d'un premier prototype de compteur réalisé par l'INRA Avignon. Dans la même thématique, [Veeraraghavan 2008] propose une chaîne d'analyse et de traitement d'images pour la détection de la fameuse danse des abeilles à partir de séquences vidéo capturées à l'intérieur de la ruche. Cela dit, la grande majorité des approches proposées dans la littérature ont pour but d'automatiser des tâches, certes laborieuses, mais déjà réalisables sans l'aide de la technologie.

Ainsi, cette thèse s'inscrit dans le cadre du développement d'outils permettant la collecte et la modélisation de données d'une nature nouvelle, jusque là inaccessible pour les biologistes. Dans cette optique, cette thèse s'inscrit donc comme une étude de faisabilité dans laquelle les verrous scientifiques et technologiques ont été identifiés et en grande partie levés. Cette étude a été menée suivant deux grandes phases.

Première phase : collecte des trajectoires de vol Faute de moyens disponibles pour les collecter à grande échelle, les trajectoires de vol des insectes n'ont été que très peu étudiées par les entomologues. Pourtant, ces trajectoires sont une source d'information potentiellement intéressante. À l'image de travaux récents proposant un dispositif multi-caméras en conditions contrôlées pour la collecte de trajectoires 3D de moustiques [Spitzen 2013], nous parions sur l'apport d'une caméra stéréovision et l'exploitation de séquences RGB-D pour suivre en 3D chaque abeille en vol devant la ruche. Ce premier objectif représente un vrai challenge étant donné la nature fortement encombrée de la scène et les conditions d'acquisition en extérieur qui sont difficilement contrôlables. La figure 1 donne un aperçu des conditions expérimentales sur le terrain ainsi que des données RGB-D recueillies.

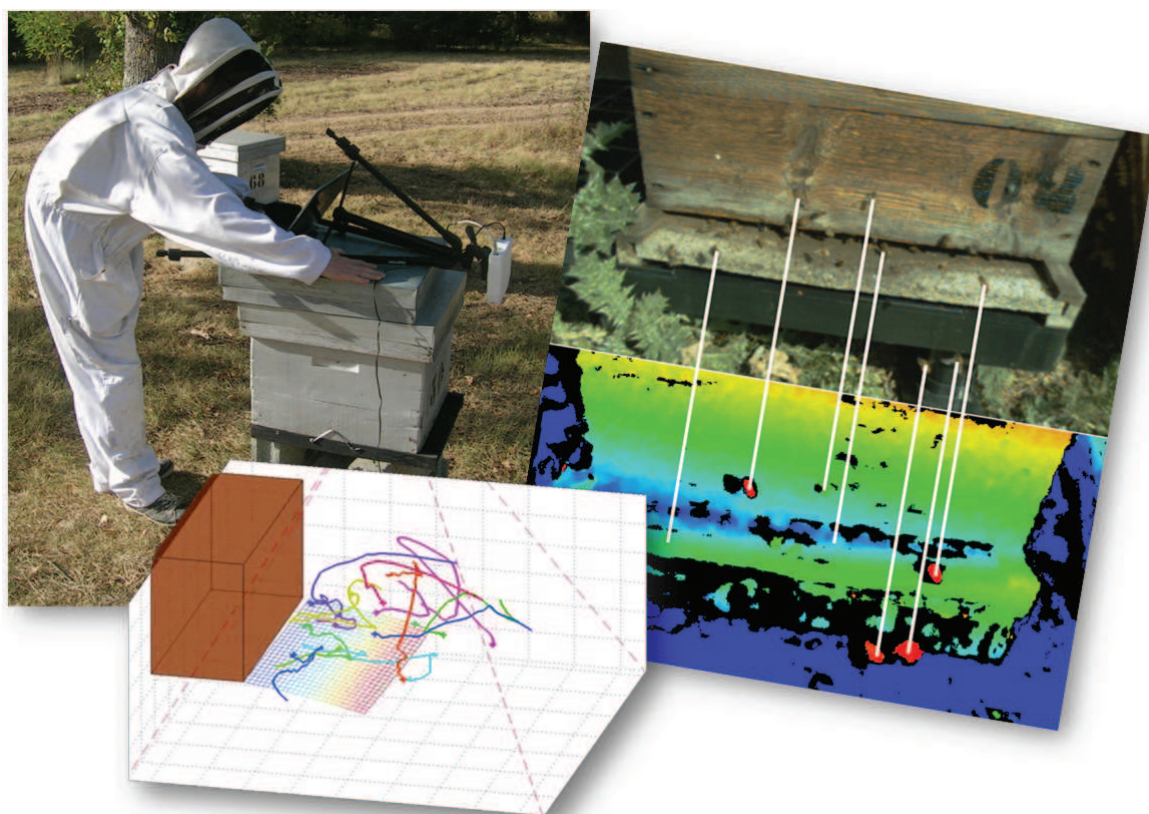


Figure 1 – À gauche, test du dispositif d'acquisition stéréovision dans le rucher du domaine du Magneraud durant l'automne 2012. À droite, couple "image RGB/carte de profondeur" capturé lors d'un autre test avec une caméra disposée devant la ruche. Les lignes blanches (tracées à la main) mettent en avant la complémentarité des deux sources d'information pour détecter les cibles en vol. En bas, reconstruction de la scène en 3D (ruche, planche d'envol et trajectoires d'abeilles).

Seconde phase : modélisation comportementale de la colonie Inspirée des différentes collaborations mentionnées précédemment, cette seconde phase vise à utiliser les trajectoires d'abeilles collectées à grande échelle pour modéliser le comportement d'une colonie et étudier son évolution dans le temps. En dépit d'un partenariat avec des collègues biologistes et apiculteurs experts, il s'avère que ce domaine d'investigation reste terra incognita pour nous tous. Ainsi, faute de pouvoir s'appuyer sur des connaissances métiers bien établies (e.g. synergie de facteurs non maîtrisée), nous sommes contraints de travailler sur des approches de modélisation comportementale nécessitant un minimum d'*a priori*. La figure 2 illustre de manière schématique le lien entre une synergie possible de facteurs susceptibles d'influencer le jeu de données de trajectoires et des comportements potentiellement identifiables.

L'étude comportementale des abeilles est une application très intéressante d'un point de vue "vision par ordinateur" et "traitement du signal". Jusqu'à aujourd'hui, le domaine de la modélisation et de l'analyse comportementale a été majoritairement dominé par des applications liées aux comportements humains, ce qui correspond à un champ d'applications finalement plutôt restreint qui peut avoir tendance à cloisonner les idées et les solutions proposées. L'étude

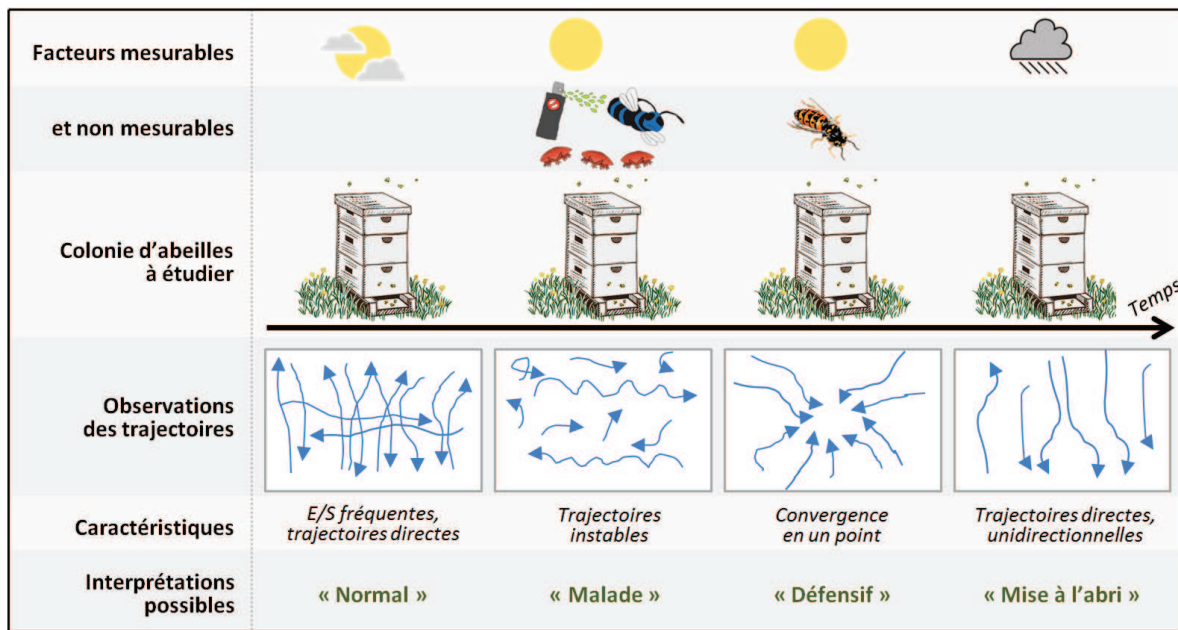


Figure 2 – Illustration schématique présentant de manière extrêmement simplifiée le problème de la modélisation comportementale en fonction des facteurs en synergie susceptibles d'influencer le jeu de données de trajectoires. Les facteurs présentés sur ce schéma sont : la météo, la présence du varroa, les traitements d'origine chimique et la prédation par le frelon asiatique.

comportementale des abeilles est finalement un bon prétexte pour explorer de nouvelles pistes dans ce domaine pour lequel les données sont particulièrement complexes (i.e. quantité, nature, connaissances métiers limitées, ...). Nous verrons dans ce manuscrit comment il a été nécessaire de réfléchir à l'ensemble de la chaîne d'analyse pour pouvoir s'adapter à la nature des données fortement contraintes.

Description du problème applicatif pour mieux apprécier les contraintes

La problématique générale de cette thèse porte sur l'étude et le développement d'une chaîne complète d'acquisition vidéo, de détection de cibles, de suivi de trajectoires, et de modélisation comportementale pour des environnements 3D de scènes encombrées en milieu naturel ; et l'application de cette dernière à la surveillance apicole. Derrière cette problématique se cachent de nombreuses contraintes que l'on peut regrouper en différentes catégories présentées maintenant.

Contraintes liées au protocole expérimental Dans le cadre de l'étude, il est nécessaire de suivre une certaine rigueur expérimentale conformément aux bonnes pratiques des biologistes. Par exemple, nous devons prendre garde à ne pas perturber la colonie lors des diverses manipulations sur le terrain (e.g. montage/démontage du dispositif d'acquisition au niveau de la ruche), ce qui modifierait le comportement des abeilles et biaiserait ainsi les données collectées.

Le protocole expérimental prévoit la capture de données issues de différents capteurs (i.e. caméra stéréovision et compteur d'entrées/sorties). Il faut donc anticiper la fusion de ces données en positionnant des repères temporels de manière à faciliter l'étape de synchronisation. Les acquisitions planifiées suivant un calendrier strict établi en amont, offre l'assurance de collecter des données sur la colonie dans ses différents états (i.e. âge des individus, cohortes introduites avec différents niveaux d'intoxication). Cela restreint cependant les marges de manœuvre en termes de tests et de prototypages.

Contraintes dues à la nature des cibles Les abeilles sont de petites cibles, rapides, évoluant de manière chaotique dans un espace à trois dimensions. À notre connaissance, il n'existe pas pour ce type de cible de modèle de dynamique à l'échelle à laquelle nous les observons. De plus, ces dernières possèdent une forte similarité entre elles, ce qui accroît les ambiguïtés de suivi lors des nombreux croisements. Paradoxalement, nous pouvons difficilement compter sur un modèle d'apparence générique pour détecter et suivre les abeilles car l'apparence de ces dernières peut varier significativement d'une image à une autre (e.g. déformation du corps, angle d'attaque, arrière plan vu par transparence au niveau des ailes...)

Contraintes d'environnement et de scène Les acquisitions sont effectuées en milieu extérieur, où les conditions d'éclairage sont difficilement maîtrisables. De plus, les nombreuses cibles en vol engendrent des ombres portées sur la planche d'envol. Ceci rend l'apparence de l'arrière plan particulièrement difficile à modéliser. Vient s'ajouter à cela le fait que nous ne souhaitons pas détecter les abeilles en déplacement (rapide ou lent) sur la planche d'envol, cela en raison de leur densité trop importante.

Contraintes liées au dispositif d'acquisition Au vu des dispositifs d'acquisition RGB-D disponibles sur le marché en 2011, le choix fut relativement limité. Outre le compromis effectué sur les caractéristiques (i.e. fréquence d'acquisition, résolution, complétude et précision des données), il a également fallu prendre en compte l'aspect physique du dispositif (i.e. dimension, niveau d'intrusion, consommation électrique, stockage des données).

Contraintes de modélisation et d'analyse Les connaissances apportées par les biologistes concernant le comportement de vol des abeilles sont limitées, ce qui rend la question de la modélisation comportementale à partir des trajectoires relativement difficile à cadrer. De plus, l'information contenue dans chaque trajectoire prise individuellement est assez pauvre, due à la dynamique chaotique des cibles. C'est pourquoi il semble plus pertinent d'étudier le comportement de la colonie dans sa globalité en s'appuyant sur des sous-ensembles de trajectoires.

Verrous scientifiques et solutions envisagées

De manière directe ou indirecte, plusieurs verrous scientifiques et technologiques ont été abordés dans la section précédente. Le tableau 1 liste les verrous que nous avons cherché à lever dans le cadre de cette thèse. Nous avons décidé de nous focaliser sur les aspects "vision par ordinateur" et "traitement du signal", ce qui permet de souligner l'importance de ces domaines dans le cadre d'une problématique applicative comme la notre. Ainsi, une ou

plusieurs solutions sont proposées pour chacun des verrous listés dans le tableau 1. L'ensemble de ces solutions sont détaillées dans ce manuscrit.

	Verrous scientifiques et technologiques	Solutions proposées
Acquisition	<ul style="list-style-type: none"> • Observation des comportements des abeilles en vol. • Collecte de données suffisamment robustes pour calculer des trajectoires individuelles. 	<ul style="list-style-type: none"> • Utilisation d'un dispositif d'acquisition vidéo positionné à l'entrée de la ruche. • Utilisation de la 3D via la caméra stéréoscopique "TYZX - G3 Evo" (RGB-D, 752 × 480 px, @50 fps).
Détection de cibles	<ul style="list-style-type: none"> • Détection et localisation des abeilles à partir de séquences RGB-D faites d'informations imprécises et incomplètes. • Absence de vérité terrain pour la validation de la méthode. 	<ul style="list-style-type: none"> • Méthode hybride de détection de cibles s'appuyant sur l'intensité et la profondeur. Méthode nommée "Segmentation HIDS". • Campagne d'annotation de données en triple aveugle.
Calcul de trajectoires	<ul style="list-style-type: none"> • Pistage 3D à partir de données incomplètes (i.e. information de profondeur manquante). • Gestion multi-cibles (croisements, apparitions/disparitions, données manquantes). • Prise en considération de la scène sur la dynamique des cibles. • Modèle de dynamique des cibles inconnu et difficilement perceptible. • Absence de vérité terrain pour la validation. 	<ul style="list-style-type: none"> • Filtre de Kalman et projection temporaire des données manquantes. • Intérêt de la 3D démontré sur les méthodes d'association de données GNN (simple) et MHT (complexe). • Méthode de suivi s'appuyant sur des hypothèses de dynamiques cibles/scène (fort <i>a priori</i>). • Méthode de suivi s'appuyant principalement sur les données (peu d'<i>a priori</i>). • Simulation de scénarios s'appuyant sur des trajectoires réelles.
Modélisation comportementale	<ul style="list-style-type: none"> • Étude comportementale à partir de trajectoires individuelles. • Modélisation avec peu d'<i>a priori</i>. • Absence de vérité terrain pour la validation. 	<ul style="list-style-type: none"> • Regroupement temporel et structuration hiérarchique des trajectoires en "captures / sousclips" pour une modélisation centrée sur la colonie dans son ensemble plutôt que sur l'individu. • Approche Bayésienne non-paramétrique s'appuyant sur le MLC-HDP. • Simulation multi-agents : modèles de colonie d'insectes (fourmis, abeilles) pour la génération de données de trajectoires et de comportements associées.
Étude de faisabilité applicative	<ul style="list-style-type: none"> • Projet EPERAS : étude des variations comportementales liées à la prédation du frelon asiatique. 	<ul style="list-style-type: none"> • Perturbations manuelles de la colonie (e.g. intrusions humaines dans le couloir de vol devant la ruche, mouvements brusques...) susceptibles de provoquer des modifications comportementales à plus ou moins long terme.

suite du tableau sur la page suivante

	<ul style="list-style-type: none"> • Projet RISQAPI : étude des variations comportementales liées au degré d'intoxication des cohortes introduites dans la colonie (abeilles équipées d'une puce RFID). 	<ul style="list-style-type: none"> • Fusion de données : 1) filtrage des données de détection d'entrées/sorties RFID via une méthode de filtrage par algorithme génétique ; 2) identification visuelle des trajectoires issues d'abeilles équipées d'une puce RFID ; 3) appariement des données d'entrées/sorties avec les trajectoires retenues.
--	--	--

Tableau 1 – Verrous scientifiques et technologiques de l'étude et aperçu des solutions proposées dans ce manuscrit.

Généricité de l'étude

Le tableau précédent met en avant le coté applicatif de cette thèse, mais celle-ci s'inscrit avant tout dans un cadre générique que l'on peut décrire au travers des trois points suivants.

Classes de problèmes étudiés Capture vidéo en milieu naturel ; appréhension 3D ; segmentation sur séquences RGB-D ; suivi multi-cibles à partir de mesures hétérogènes ; analyse de scènes encombrées ; recherche de motifs temporels récurrents.

Point de vue méthodologique adopté Valider indépendamment chaque étape de la chaîne à l'aide d'une vérité terrain dédiée ; minimiser l'*a priori* au niveau des différentes approches proposées (i.e. suivi de cibles, modélisation comportementale, simulation de trajectoires) permettant de limiter la dépendance à des connaissances métiers éventuellement absentes ou non maîtrisées.

Algorithmes génériques proposés La segmentation HIDS peut être adaptée à d'autres problèmes de détection de cibles à partir de séquences RGB-D ; les solutions et améliorations présentées pour le calcul des trajectoires, avec ou sans *a priori*, peuvent répondre à un ensemble de problèmes mettant en œuvre un grand nombre de cibles ; cette thèse contribue à démontrer la généricité des approches Bayésiennes non-paramétriques, et en particulier le MLC-HDP (originellement proposé dans le cadre d'une application médicale [Wulsin 2012]). Ces méthodes font partie des innovations récentes concernant la modélisation de données structurées avec peu d'*a priori*.

Organisation du manuscrit

La figure 3 donne un aperçu général du plan de la thèse et des objectifs de chaque chapitre présenté. Les chapitres 1, 2 et 3 abordent successivement les différents maillons de notre chaîne d'acquisition et de traitement de données. Le chapitre 4, quant à lui, détaille la mise en œuvre de la chaîne complète sur des données réelles, notamment au travers des collaborations avec l'INRA Magneraud et le Muséum National d'Histoire Naturelle dans le cadre des projets RISQAPI et EPERAS. Chaque élément de la chaîne est optimisé pour atteindre in fine notre objectif de modélisation comportementale qui fait donc office de fil rouge tout au long de ce manuscrit.

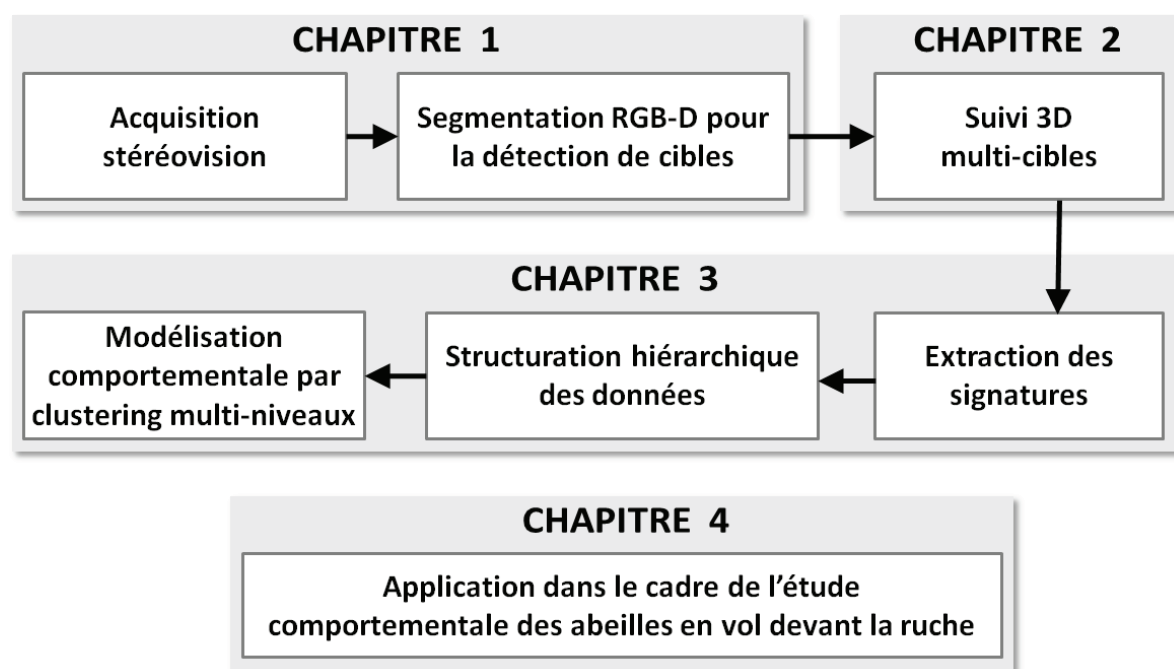


Figure 3 – Découpage en 4 chapitres de notre chaîne d'acquisition et de traitement de données.

Chaque chapitre aborde un ou plusieurs maillons de la chaîne et se focalise sur une problématique particulière. Ainsi, les chapitres, en fonction du domaine d'intérêt du lecteur, peuvent être lus de manière quasi-indépendante. Tous les chapitres sont construits selon une structure commune, composée :

- **d'une introduction** destinée à replacer le contexte du chapitre, présenter la problématique abordée, synthétiser en un bref résumé l'état de l'art, et aborder les solutions/contributions proposées ;
- **d'un état de l'art** aussi exhaustif que possible sur le domaine concerné par le chapitre ;
- **des solutions proposées** qui sont pour une partie d'entre elles innovantes vis-à-vis des travaux mentionnés dans l'état de l'art. Certaines ont d'ailleurs fait l'objet de publications dans des conférences et des journaux internationaux ;
- **d'une évaluation** des solutions proposées à l'aide de données réelles ou simulées.

Le lecteur pourra apprécier l'ensemble des contributions apportées par cette thèse dans la conclusion générale. De même, la liste des publications relatives à nos travaux est donnée à la fin de ce manuscrit.

Résumé des chapitres

Chapitre 1 En premier lieu, dans ce chapitre, nous proposons d'étudier les contraintes inhérentes à la collecte de trajectoires d'abeilles devant la ruche. Suite au constat d'un encombrement potentiellement fort de la scène, le recours à l'information de profondeur est

envisagé. Nous comptons sur une appréhension 3D de la scène pour lever les ambiguïtés liées aux occultations et aux ombres portées. Ainsi, à l'issue d'une étude sur les principaux dispositifs d'acquisition RGB-D disponibles sur le marché en 2011, nous avons décidé d'adopter la stéréovision à l'aide de la caméra TYZX G3 Evo.

Une seconde difficulté est liée à la faible dimension des abeilles à détecter. Nous avons proposé une méthode de détection innovante que nous nommons "Segmentation HIDS". Cette méthode s'inspire des approches de détection par soustraction d'arrière plan, mais ajoute un mécanisme de fusion d'information permettant de tirer le meilleur parti des informations contenues dans chaque couple "image d'intensité/carte de profondeur" de la séquence vidéo. Lorsque c'est possible, les cibles détectées sont localisées en 3D (i, j en pixels dans le repère image, et d en mm), mais lorsque l'information de profondeur n'est pas disponible ou n'est pas fiable, on se contente d'une définition 2D (i, j).

Pour clore ce premier chapitre, nous proposons finalement d'évaluer notre segmentation HIDS sur des scènes réelles capturées devant la ruche.

Chapitre 2 Ce second chapitre s'attaque au calcul des trajectoires 3D. Il s'agit de trouver une méthode permettant de suivre de nombreuses cibles (jusqu'à 20) simultanément sans que les trajectoires se confondent les unes avec les autres. L'analyse 3D rend plus facile la phase de mise en correspondance traces/observations. Ainsi nous verrons qu'il n'est pas forcément nécessaire d'avoir recours à des méthodes complexes et gourmandes en temps de calcul. Nous proposons d'utiliser une approche s'appuyant sur le couple "Filtre de Kalman / Global Nearest Neighbor". Notons que pour des raisons pratiques, nous exprimons les trajectoires dans un espace 3D défini par des grandeurs réelles (échelle sur les axes x , y et z exprimée en cm). Ceci nécessite, pour chaque cible détectée, d'appliquer une projection à l'aide des paramètres intrinsèques de la caméra.

La méthode de segmentation HIDS, aussi robuste soit-elle, renvoie parfois des données 2D (dépourvues de profondeur), ce qui est incompatible avec la solution de suivi choisie initialement. Ainsi, nous proposons de traiter ces situations grâce à une estimation anticipée de la profondeur lors de la phase de projection dans l'espace de suivi 3D. Dans ce chapitre, nous proposons également deux méthodes de suivi de cibles distinctes s'appuyant sur des approches fondamentalement opposées, pour la prise en compte de l'influence de la scène sur la dynamique des cibles.

Les différentes méthodes proposées sont évaluées grâce à une étude comparative reposant sur des scénarios semi-simulés mettant en œuvre un nombre plus ou moins important de trajectoires.

Chapitre 3 Dans ce troisième chapitre, nous étudions le problème de la modélisation comportementale d'une colonie à partir des trajectoires des individus. Confrontés au manque de connaissance métiers disponibles sur les trajectoires de vol d'abeilles, nous jugeons opportun d'adopter une méthode de modélisation par approche Bayésienne dite "Non-paramétrique". Dans ce cadre, nous proposons d'adapter à notre problème le MLC-HDP (pour Hierarchical Dirichlet Process Model with Multiple Levels of Clustering). Le MLC-HDP originalement proposé par Drausin Wulsin en 2012 dans le cadre d'une application médicale, est une méthode de clustering agissant simultanément sur les différents niveaux d'une structure de données hiérarchisée.

Ainsi, dans ce chapitre, nous proposons notre propre structure selon des entités temporelles relatives aux acquisitions (i.e. capture et sousclips). Celle-ci permet de regrouper les trajectoires et de condenser l'information contenue dans chacune de celle-ci, afin de créer des ensembles porteurs de sens et ainsi exploitables. Nous proposons également notre propre signature permettant de représenter une trajectoire 3D en un vecteur de quelques valeurs seulement.

À notre connaissance, il n'existe pas à ce jour de vérité terrain faisant le lien entre les trajectoires et les comportements d'une colonie d'insectes. Nous proposons ainsi dans ce chapitre une méthode originale pour valider notre approche. Celle-ci repose sur l'utilisation de Système Multi-Agents (SMA) via des modèles de simulation d'une colonie d'insectes (fourmis ou abeilles) capables de générer des comportements et des trajectoires associées.

Chapitre 4 Ce quatrième et dernier chapitre s'intègre directement dans la cadre des protocoles expérimentaux mis en place pour les projets EPERAS et RISQAPI. Ceci se traduit par la mise en application à grande échelle de notre chaîne d'acquisition et de traitement sur des données réelles issues d'acquisitions sur le terrain. L'étude comportementale des abeilles à partir de leurs trajectoires de vol devant la ruche est totalement novatrice, d'un point de vue applicatif, mais également d'un point de vue méthodologique. En effet, les données collectées sont uniques en leur genre (i.e. quantité, nature, connaissances métiers disponibles, ...). Ce chapitre montre également comment, l'implication forte dans un projet de recherche interdisciplinaire, nous a conduit à réfléchir et à développer par rapport à notre dispositif initial, des briques supplémentaires (méthode de détection visuelle de puces RFID, méthode de nettoyage des données d'entrées/sorties collectées par le dispositif de comptage RFID, méthode de fusion de données). Notons que ces briques sont encore aujourd'hui à l'état de prototype et que la chaîne complète permettant d'exploiter pleinement les données biologiques est aujourd'hui encore en cours de d'introspection.

Acquisition stéréovision et détection de cibles

Sommaire

1.1	Introduction	16
1.2	Dispositif d'acquisition	16
1.2.1	Contraintes applicatives	17
1.2.2	Choix du système d'acquisition 3D	18
1.2.3	Configuration du dispositif stéréovision	24
1.3	État de l'art sur la détection de cibles	25
1.3.1	Détection par modèle d'apparence	25
1.3.2	Détection par le mouvement	26
1.3.3	Détection sur des données RGB-D	27
1.3.4	Détecteurs utilisés pour les insectes	27
1.4	Solution proposée : une segmentation hybride	28
1.4.1	Complémentarité des données d'intensité et de profondeur	28
1.4.2	Algorithme HIDS (Hybrid Intensity Depth Segmentation)	31
1.4.3	Extraction des cibles	37
1.5	Évaluation de la méthode de segmentation HIDS	38
1.5.1	Élaboration de la vérité terrain	39
1.5.2	Résultats	41
1.5.3	Synthèse sur l'influence des paramètres	43
1.6	Conclusion	44

1.1 Introduction

Ce premier chapitre pose les bases d'un dispositif de vidéosurveillance fonctionnant en extérieur, pour la capture et l'appréhension 3D de scènes encombrées. Le système est conçu pour observer, détecter et localiser un nombre élevé de petites cibles évoluant dans un volume ouvert selon une dynamique rapide et erratique. Les premiers tests d'acquisition sur le terrain ont permis de mesurer les difficultés inhérentes à l'observation du vivant dans son milieu naturel. Ainsi, ce chapitre propose d'étudier les contraintes et de lever les verrous scientifiques et technologiques typiques de ce domaine (e.g. suivi d'abeilles devant la ruche).

Problématique On s'intéresse ainsi à la double problématique suivante : *"quels matériels et méthodes pour la détection de trajectoires individuelles de petites cibles en évolution rapide ?"*. Tout d'abord, il est essentiel de bien cerner les contraintes applicatives (e.g. conditions d'acquisition, type de scène, nature des cibles) afin de choisir en conséquence un système d'acquisition adéquat. Un compromis devra être trouvé, car aucun produit proposé sur le marché au moment de l'étude ne répond parfaitement aux exigences. Puis, en fonction de la nature des données collectées, des méthodes de détection de cibles existantes, voire nouvelles devront être mises en œuvre. Notons que ces premiers choix matériels et méthodologiques sont capitaux, car ils constituent le premier maillon d'une chaîne d'acquisition et de traitement sur lequel d'autres traitements vont s'appuyer, notamment le calcul des trajectoires qui sera abordé dans le chapitre suivant.

Littérature La littérature regorge de travaux sur la détection de cibles à partir de séquences ou d'images 2D (cf. étude comparative [Sobral 2014]). En revanche, la démocratisation des dispositifs d'acquisition 3D (notamment RGB-D) — caméras stéréoscopiques, caméras temps-de-vol, caméras 3D par projection de lumière structurée, caméras plénoptiques — étant relativement récente, la littérature concernant la détection de cibles en 3D utilisant ce type de matériel est plutôt restreinte.

Solution proposée Du côté matériel, nous avons finalement choisi d'utiliser une caméra stéréoscopique (dite aussi stéréovision) pour capturer des séquences RGB-D de la scène, soit le modèle G3 Evo produit par la société TYZX. Du côté traitement d'images, pour la détection de cibles en 3D, nous proposons une méthode innovante de segmentation (que nous avons appelée "Segmentation HIDS") s'appuyant sur une triple fusion d'information (deux sources d'information liées au mouvement, et une troisième liée à la profondeur). Cette approche semble suffisamment robuste pour employer par la suite une méthode de suivi de cibles de type "detect-before-track" (cf. Chapitre 2).

1.2 Dispositif d'acquisition

Rappelons-le, l'étude comportementale des abeilles est le fil rouge que nous allons suivre tout au long de ce manuscrit. En premier lieu, il est essentiel de bien identifier les contraintes liées à l'application afin de choisir en conséquence un système d'acquisition adéquat.

1.2.1 Contraintes applicatives

Nous proposons dans cette section de partager l'expérience acquise sur le terrain lors de tests préliminaires d'acquisition de séquences vidéo devant des ruches. Voici la liste des contraintes fortes liées à l'acquisition et au traitement d'image.

Nombre élevé de cibles La figure 1.1 montre une activité importante devant la ruche. Les résultats de comptage obtenus par [Blois 2011] montrent que les abeilles peuvent aller et venir dans la ruche par vagues relativement denses. Dans ces conditions, les nombreuses occultations peuvent rendre la détection et le suivi des cibles particulièrement difficiles.

Apparence des cibles Les abeilles sont toutes similaires, mais paradoxalement leur apparence sur les images d'intensité (ou couleur) peut grandement varier en fonction de l'action en cours (e.g. vol, ventilation de la ruche, déplacement sur la planche d'envol, ...), de leur position, de leur orientation, et de l'éclairage de la scène.

Taille des cibles Une abeille mesure environ 1.2 cm sur 0.6 cm , ce qui en fait une petite cible relativement à l'échelle de la scène observée. Les abeilles évoluent dans un espace à 3 dimensions et peuvent ainsi se trouver à une distance variable du capteur, ce qui amène à réfléchir aux limites de portée d'observation de celui-ci. Pour assurer une fonctionnalité de comptage exhaustive des entrées/sorties, il est nécessaire de capturer la planche d'envol et en particulier l'entrée de la ruche dans sa totalité. La largeur de la planche d'envol est de 50 cm . En termes d'acquisition, il faut donc trouver un compromis entre le champs de vision, la portée, la précision/résolution du capteur qui satisfasse l'ensemble de ces contraintes.

Dynamique des cibles Le mouvement des abeilles est difficilement prédictible. Elles peuvent voler à des vitesses allant jusqu'à 8 m/s , et ainsi traverser le champ de la caméra en quelques dixièmes de seconde, ce qui couvre un nombre relativement faible d'images si l'on considère une fréquence d'acquisition moyenne (e.g. 24 Hz). Même si la plupart des abeilles ralentissent à l'approche de la planche d'envol, leurs changements soudains de direction imposent l'utilisation d'un dispositif d'acquisition haute fréquence. Dans notre application, nous ne nous intéressons cependant pas au battement des ailes des abeilles, qui pour être étudié, nécessiterait un dispositif d'acquisition avec un temps d'exposition très court pour éviter le flou lié au battement très rapide.

Éclairage de la scène Les acquisitions sont effectuées en extérieur, dans le milieu naturel des abeilles, où les conditions d'éclairage sont difficilement maîtrisables. Ainsi, en fonction de l'activité devant la ruche, les images capturées peuvent contenir parfois plus d'ombres projetées sur la planche d'envol que d'abeilles réellement présentes dans le champ de la caméra. L'exposition directe à la lumière du soleil rend l'utilisation des dispositifs infrarouges peu recommandée (e.g. Microsoft Kinect).

Arrière plan Les auteurs de [Campbell 2008] (travaux abordés plus tard dans ce manuscrit) détectent les abeilles sur une planche d'envol propre et de couleur claire, ce qui semble être le cas optimal. Cependant, dans la plupart des cas, la planche d'envol se salit rapidement avec

le temps. Et celle-ci n'est par ailleurs pas forcément blanche à l'origine. Parfois même, il est difficile de distinguer les abeilles de la planche à l'œil nu à cause des salissures.

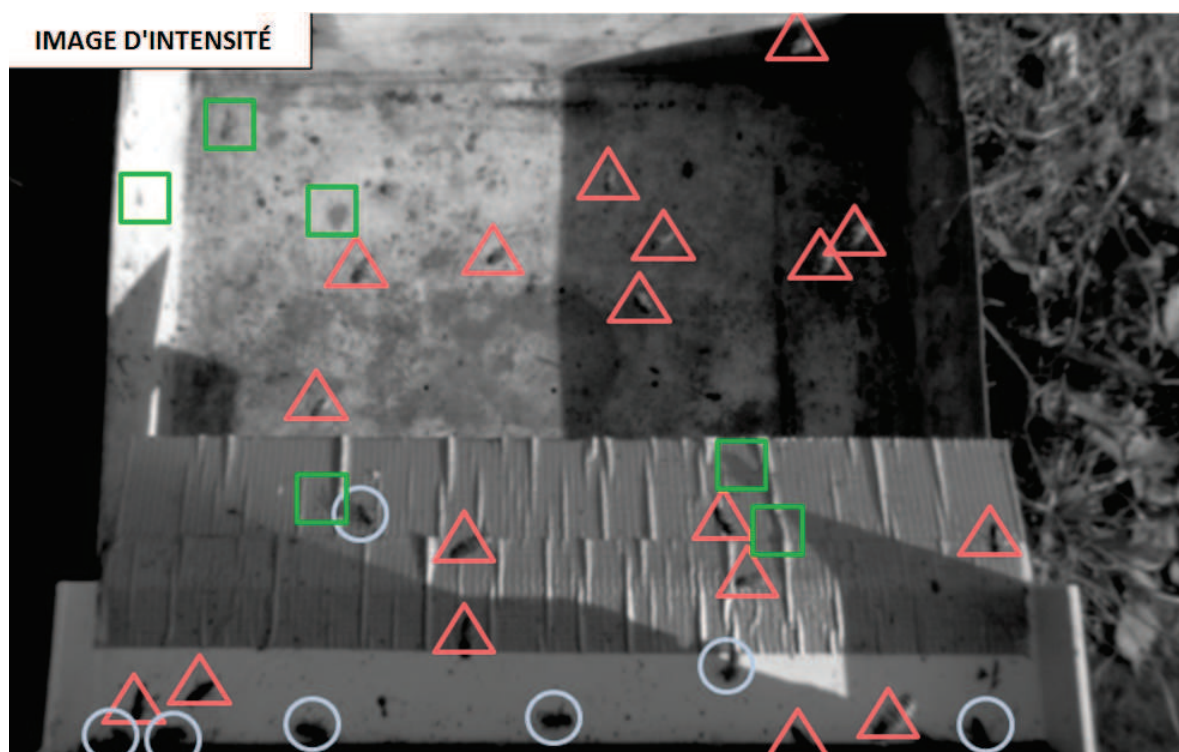


Figure 1.1 – Image d'intensité montrant une activité moyenne des vols à l'entrée d'une ruche. Les marques ont été annotées manuellement à l'aide d'un logiciel de dessin. Les abeilles ont pu être distinguées des ombres portées grâce au mouvement observé à partir de la séquence. Ainsi, on note 19 abeilles en vol (triangles rouges), 7 abeilles en déplacement sur la planche d'envol (cercles bleus) et 6 ombres projetées d'abeilles (carrés verts). La séquence vidéo a été acquise lors des tests sur le terrain effectués en collaboration avec l'INRA Magneraud durant l'été 2011.

Compte tenu des tests terrain et des contraintes mentionnées ci-dessus, nous avons décidé d'adopter une capture 3D de la scène afin de détecter correctement les insectes en vol dans leur milieu naturel. Ainsi, la section 1.2.2 fait un point sur les dispositifs d'acquisition 3D disponibles sur le marché en 2011. Nous verrons pour quelles raisons notre choix s'est porté finalement sur la technologie stéréovision, et décrirons ensuite la configuration choisie (cf. Section 1.2.3).

1.2.2 Choix du système d'acquisition 3D

1.2.2.1 Technologies existantes

Tel que mentionné en introduction, en 2011, l'année durant laquelle nous avons lancé cette étude, il n'existait pas de dispositif d'acquisition parfaitement adapté à notre application. Voici une liste actualisée des principales technologies existantes en termes d'acquisition 3D :

Multi-caméras La scène est observée simultanément à partir de différents points de vue par des capteurs distincts. [Wang 2013] est un résumé récent des différentes méthodes existantes. Le calcul des informations 3D se fait par fusion d'information, éventuellement à différents niveaux :

- données brutes, par correspondance de points d'intérêts entre des paires de caméras. (e.g. [Dockstader 2001]) ;
- données de cibles, par triangulation (e.g. [Straw 2010, Theriault 2010]) ou correspondance d'apparence (e.g. [Zhou 2006]) ;
- données de trajectoires, par correspondance de forme (e.g. [Nummiaro 2003, Zhu 2011]).

Exemple de produit commercial : Optitrack

Stéréovision S'appuie également sur le principe de fusion de données brutes. La scène est observée depuis 2 ou 3 points de vue légèrement décalés. Avec la connaissance des paramètres de calibrations intrinsèques (interne à chaque capteur), et extrinsèques (inter-capteurs), une carte de profondeur (aussi appelée carte des disparités) peut être reconstruite à l'issue d'une étape d'appariement de points d'intérêts. Les caméras stéréovision ont l'avantage d'offrir une solution d'acquisition 3D prête à l'emploi.

Exemple de produits commerciaux : Point Grey Bumble Bee 2, TYZX G3 EV

Temps de vol Les caméras Time of Flight (TOF) mesurent l'intensité et le temps de vol d'un signal lumineux (généralement infrarouge) entre l'objectif et la scène pour chaque point de l'image. L'information de distance est calculée par rapport à la vitesse de la lumière.

Exemple de produits commerciaux : Microsoft Kinect v2, PmdTec Camboard Nano

Téledétection par laser Les LIDARs (pour Light Detection and Ranging) fonctionnent sur le principe de temps de vol de la lumière, à la différence que l'acquisition de la scène se fait de manière progressive par un unique laser amovible. La résolution est bien meilleure au détriment de la vitesse d'acquisition.

Exemple de produit commercial : Hokuyo 2D Laser Scanner

Lumière structurée Une grille structurée est projetée (généralement en infra-rouge) sur la scène. Puis un ou plusieurs capteurs viennent observer les déformations de la grille témoignant de la topologie de la scène (cf. [Scharstein 2003]).

Exemples de produits commerciaux : Leap motion, Microsoft Kinect v1

Plénoptique Aussi appelé "Light-field camera", la technologie plénoptique emploie une matrice de micro-objectifs permettant d'enregistrer les informations dites "4D" d'un flux lumineux. L'idée est de capturer pour chaque rayon lumineux une paire de points 2D à l'aide de deux capteurs positionnés l'un derrière l'autre avec un léger décalage. Cela permet de retrouver l'angle des rayons lumineux. Des traitements permettent ensuite de calculer des informations 3D sur la scène à l'aide de l'ensemble des informations 4D capturées (cf. [Adelson 1992]).

Exemple de produit commercial : Lytro

Concernant le choix de la technologie, il s'agit maintenant de trouver un compromis par rapport aux caractéristiques espérées vis-à-vis de nos contraintes, à savoir : une haute fréquence d'observation, une résolution élevée, un champ de vision relativement grand, et une précision correcte au niveau des données recueillies. Aussi, dans l'idéal, le système d'acquisition doit être complètement non-intrusif. Notons à l'occasion que les abeilles ne sont pas sensibles à la lumière infrarouge [Von Frisch 1964].

1.2.2.2 Produits candidats

Nous portons dans cette section notre attention sur deux technologies d'acquisition 3D (aussi appelées capteurs RGB-D) : les caméras TOF et les caméras stéréovisions. Contrairement aux dispositifs multi-caméras, les caméras stéréovisions "tout en un" ont l'avantage d'être déjà calibrées en usine et fournissent directement, comme les caméras TOF, des séquences RGB-D. Le terme RGB-D fait référence au couple "image couleur (ou niveaux de gris) / carte de profondeur". Le tableau 1.1 compare les spécifications de 6 caméras proposées par les principaux fabricants de systèmes temps de vol et stéréovision en 2011.

	Camera	Spécification	Fournisseur
Stéréovision	Bumblebee 2	640 × 480 px 48 fps	Point Grey Research (Richmond, Canada)
	G3 Evo	752 × 480 px 50 fps	TYZX (Menlo Park, CA, USA)
	SVC 2	752 × 480 px 30 fps	Focus Robotics (Hudson, NY, USA)
Temps de vol	CamBoard nano	160 × 120 px 90 fps	Pmdtec (Dresde, Allemagne)
	D70	160 × 120 px 50 fps	TYZX (Stockholm, Suède)
	SR4000	176 × 144 px 50 fps	MESA Imaging (Zurich, Suisse)

Tableau 1.1 – Comparaison des spécifications (résolution et fréquence d'acquisition) de 6 caméras disponibles chez les principaux fabricants de systèmes temps de vol et stéréovision en 2011.

Caméras TOF Les spécifications des dispositifs TOF présentées dans le tableau 1.1 (en particulier les résolutions) sont trop faibles pour satisfaire les contraintes de notre application. Néanmoins, la piste TOF a tout de même retenue notre attention car elles pouvaient s'avérer complémentaire. C'est pourquoi nous avons décidé d'entreprendre des tests avec la caméra CamBoard Nano. En dépit de sa faible résolution, cette caméra offre une fréquence d'acquisition très élevée, ce qui est un atout majeur pour améliorer le suivi de cibles rapide telles que nos abeilles en vol. Nous avons d'ailleurs envisagé, sans concrétiser l'idée, de coupler plusieurs caméras de ce type pour couvrir l'intégralité de la scène à une résolution suffisante.

Caméras stéréovisions Nous avons également choisi de faire des tests sur le modèle G3 Evo qui semblait correspondre au mieux aux contraintes de notre application. En effet, ce modèle annonce une fréquence d'acquisition plus faible (60 Hz) mais se hisse tout de même au sommet de sa catégorie grâce au déport des calculs des disparités sur un processeur embarqué.

1.2.2.3 Tests : CamBoard Nano et G3 Evo

La figure 1.2 montre en photo les deux caméras CamBoard Nano (TOF) et G3 Evo (stéréovision) utilisées pour nos tests.

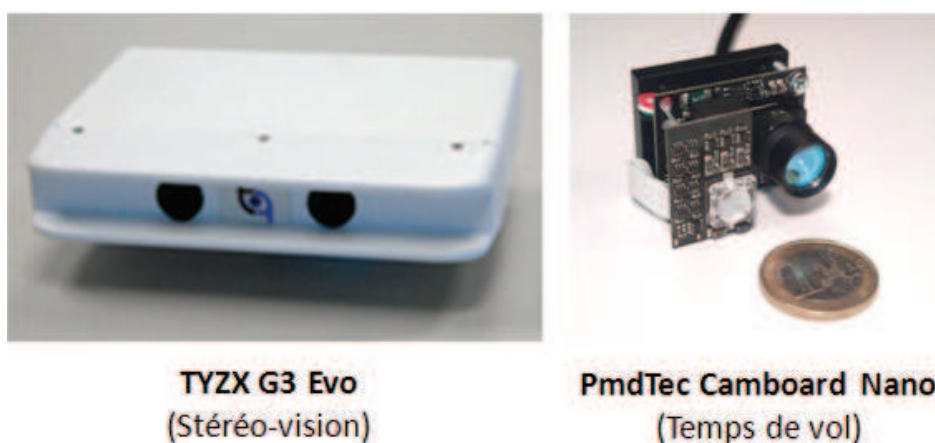


Figure 1.2 – Photos des deux caméras CamBoard Nano (TOF) et G3 Evo (Stéréovision) utilisées pour nos tests.

La figure 1.3 montre pour une scène d'entrée de ruche les couples "image d'intensité / carte de profondeur" capturés par les 2 caméras "G3 Evo" et "CamBoard Nano".

TYZX G3 Evo

En termes d'intensité lumineuse, l'image a) de la figure 1.3 montre que la scène est clairement capturée en couleur (RGB) dans sa globalité. Notons que le second objectif de la caméra capture la scène en niveaux de gris, et que l'appariement se fait à partir d'informations en niveaux de gris.

En termes d'information de profondeur, on remarque sur l'image c) des zones d'ombres (régions noires) dues à un appariement non effectué entre les 2 images capturées. Trois causes majeures de non-appariabilité ont été identifiées :

1. une texture trop faiblement contrastée (e.g. surface homogène ou un flou important provoqué par un mouvement trop rapide),
2. une différence au niveau de l'apparence d'une même zone selon le point de vue,
3. des zones en dehors de la portée par rapport aux seuils prévus par l'algorithme d'appariement.

La figure 1.4 illustre de manière schématique les problèmes 2 et 3.

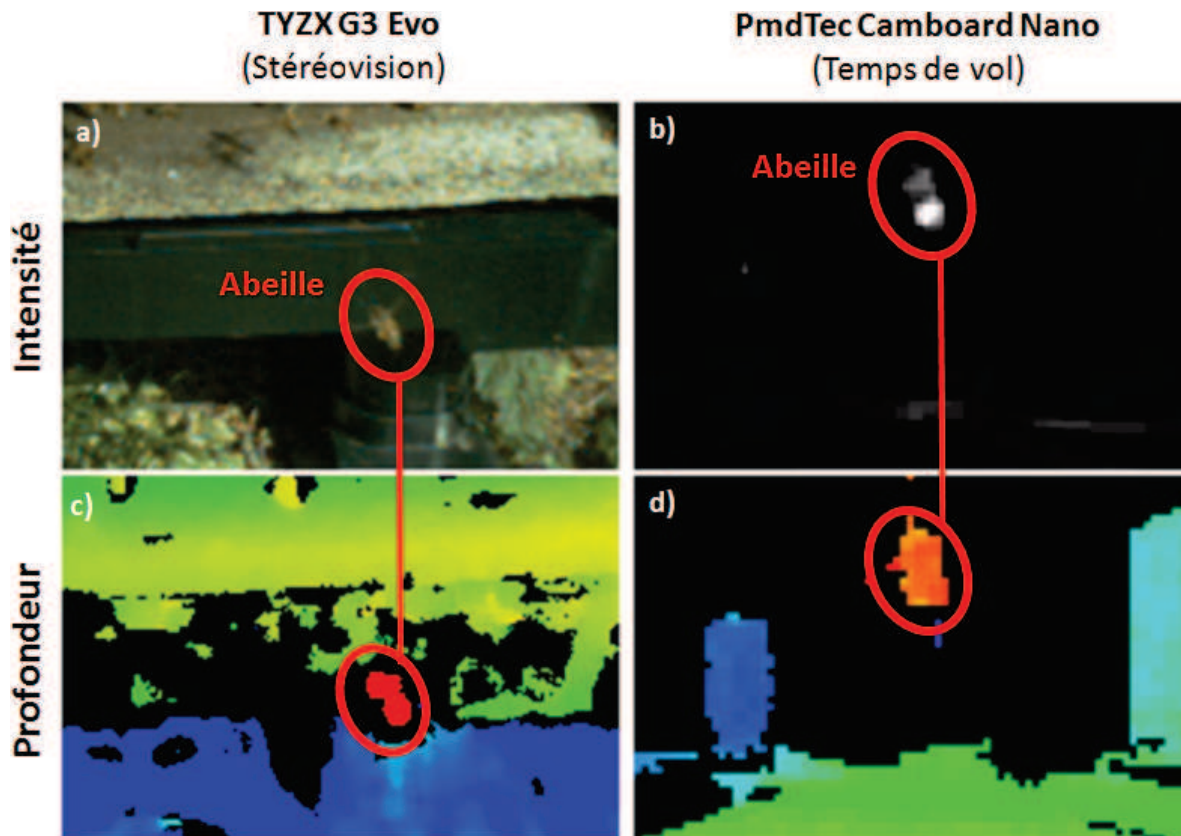


Figure 1.3 – Comparaison des données acquises par 2 caméras RGB-D issues de technologies différentes : "TYZX G3 Evo" (Stéréovision) et "PmdTec CamBoard Nano" (Temps de vol). Les images en haut correspondent aux images d'intensité lumineuse (en couleur pour la stéréovision), et celles en bas correspondent à la carte de profondeur (ou carte des disparités pour la stéréovision).

PmdTec CamBoard Nano

En termes d'intensité lumineuse, l'image b) montre que la scène est représentée avec des contrastes très prononcés. Il est d'ailleurs difficile de reconnaître l'environnement et les cibles. Au delà d'une certaine distance (environ 40 cm), le capteur TOF ne capte presque plus de lumière. On remarque également que la caméra TOF a tendance à capturer plus de lumière au centre qu'en périphérie.

Tout comme l'intensité, la portée restreinte de la caméra conduit à des zones lointaines dépourvues d'information de profondeur. On note également que les zones de la scène qui reflètent mal la lumière posent également problème.

En conclusion, le tableau 1.2 résume les forces et les faiblesses des deux caméras testées vis-à-vis des contraintes de notre application. À partir de là, nous choisissons de mener la suite notre étude avec la caméra TYZX G3 Evo qui offre visiblement le meilleur compromis résolution/fréquence d'acquisition. Les lecteurs qui souhaitent plus de détails sur les caméras temps de vol peuvent se référer à [Piatti 2011].

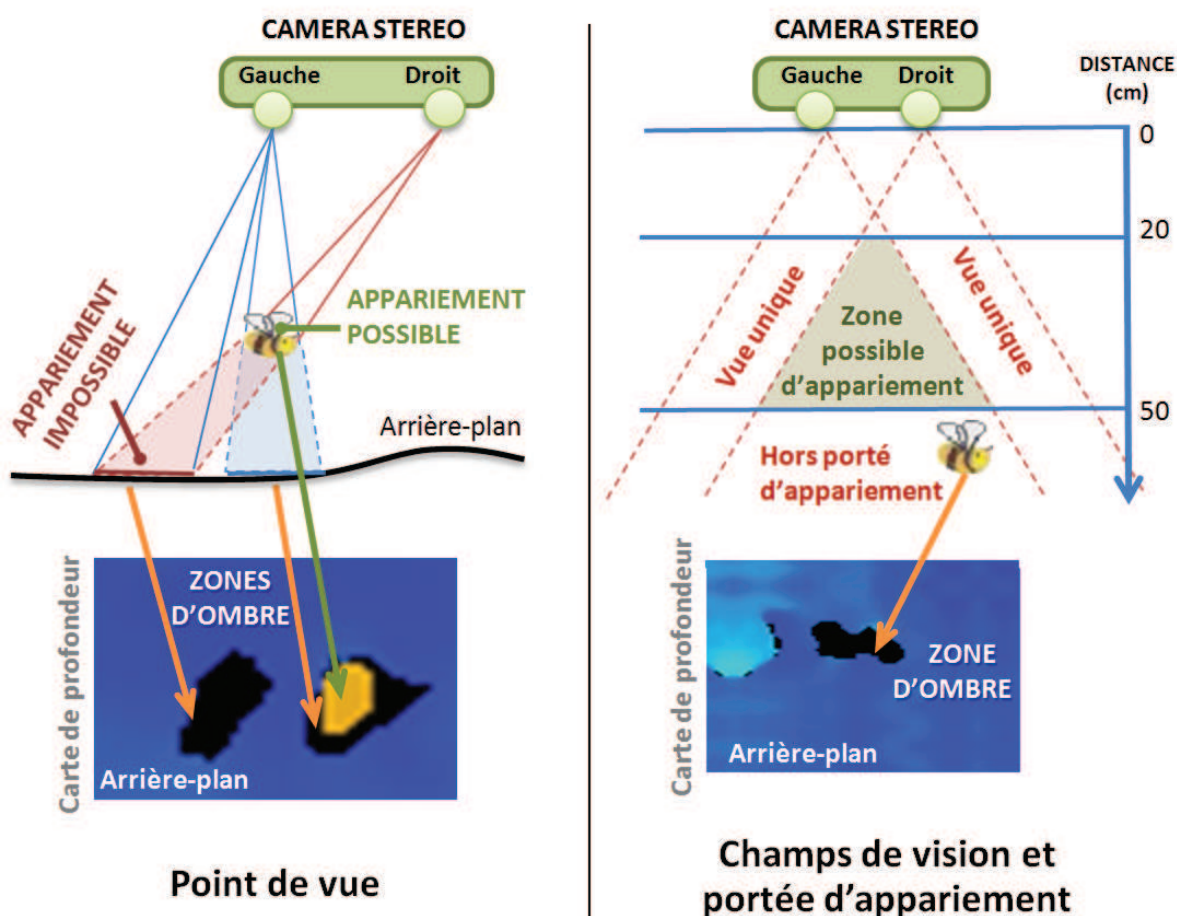


Figure 1.4 – Illustration des phénomènes de "zones d'ombre" observées sur les cartes de disparité fournies par la caméra stéréovision. À droite : problème d'occultation de l'arrière plan dû aux points de vue différents. À gauche : problème de champ de vision stéréo et portée d'appariement (lié au seuil de disparité maximum entre deux points).

	G3 Evo (Stéréovision)	Camboard Nano (Temps de vol)
Fréquence d'acquisition	++	+++
Résolution d'image	+++	+
Détection de petites cibles	+++	+++
Détection de cibles rapides	++	+++
Précision de la carte de profondeur	+++	++
Disponibilité de l'information de profondeur	++	+++

Tableau 1.2 – Résumé des forces et faiblesses des 2 systèmes "TOF CamBoard Nano" et "Stéréovision G3 Evo" relativement à nos contraintes applicatives.

1.2.3 Configuration du dispositif stéréovision

Nous l'avons vu précédemment, la résolution horizontale (H_{res}) de la caméra stéréovision TYZX G3 Evo est de 752 pixels. Le constructeur propose 3 lentilles offrant des angles de champs horizontaux (HFOV pour Horizontal Field Of View) de 40°, 62° et 80°. Également, il est possible de choisir parmi 4 configurations d'écart inter-objectifs (ou baseline), soit 3 cm, 6 cm, 8 cm et 14 cm. Le tableau 1.3 donne quelques chiffres quant au domaine d'utilisation de la caméra concernant la profondeur Z .

↓ Baseline	Lentilles →	40° HFOV ($Z \pm \Delta Z$)	62° HFOV ($Z \pm \Delta Z$)	80° HFOV ($Z \pm \Delta Z$)
3 cm	Min Z	48.4 ± 0.06	29.3 ± 0.04	21.0 ± 0.03
	Max Z	762 ± 14.4	462 ± 8.70	331 ± 6.20
6 cm	Min Z	96.8 ± 0.12	58.6 ± 0.07	42.0 ± 0.05
	Max Z	1525 ± 28.8	923 ± 17.4	661 ± 15.5
8 cm	Min Z	129 ± 0.16	78.2 ± 0.09	56.0 ± 0.07
	Max Z	2033 ± 38.4	1232 ± 23.2	882 ± 16.6
14 cm	Min Z	226 ± 0.28	137 ± 0.17	96.0 ± 0.12
	Max Z	3558 ± 67.1	2155 ± 40.7	1543 ± 29.1

Tableau 1.3 – Domaine d'utilisation de la caméra stéréovision G3 Evo concernant la profondeur Z (min. et max. et erreurs relatives), en fonction des différentes configurations proposées par le constructeur TYZX. Chiffres donnés en centimètres.

Contrainte d'observation de la scène Dans notre application, il est nécessaire d'observer la planche d'envol (mesurant 50 cm de large) dans sa totalité. Pour cela, on peut jouer sur le choix des lentilles et la distance de la caméra par rapport à la zone observée. Plus l'angle de champ horizontal est petit, plus il est nécessaire d'éloigner la caméra. Les contraintes ci-dessous vont nous aider à trouver la solution optimale.

Contrainte de détection des cibles On peut calculer la limite de profondeur au delà de laquelle une cible de type abeille (mesurant 1.2 cm par 0.6 cm) ne peut plus être détectée de manière satisfaisante. À savoir que nous avons fixé arbitrairement à 8 pixels (sur la dimension la plus petite) le seuil au-dessous duquel on considère qu'il n'est plus possible de détecter correctement une cible en 3D. Voici comment calculer le nombre de pixels par centimètre (ppc) à une profondeur donnée :

$$ppc = \frac{H_{res}}{2h \tan\left(\frac{\alpha}{2}\right)} \quad (1.1)$$

avec H_{res} la résolution horizontale de la caméra, α l'angle de champ horizontal (HFOV) et h la profondeur (distance par rapport à la caméra).

Maximisation de l'espace d'observation Tout en respectant les contraintes ci-dessus, on souhaite observer les cibles dans un volume le plus important possible, de manière à maximiser la quantité d'information collectée. Ainsi, la meilleure configuration pour notre application

s'avère être la suivante : la caméra stéréovision TYZX G3 Evo placée à 50 cm au dessus de la planche d'envol, avec un écart inter-objectifs de 3 cm et une lentille offrant un champ de vision horizontal de 62° . À l'issue de tests, nous avons constaté que la caméra peut capturer la scène à une fréquence moyenne de 47 Hz (variations situées entre 25 et 53 Hz). La figure 1.5 illustre la configuration de notre dispositif d'acquisition positionné sur une ruche.

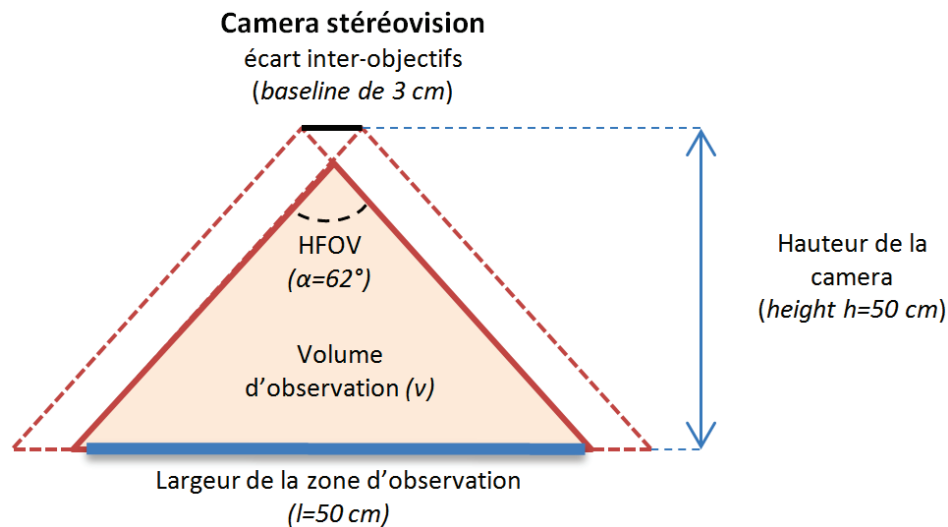


Figure 1.5 – Configuration du dispositif d'acquisition stéréovision.

1.3 État de l'art sur la détection de cibles

L'expression "détection de cibles" fait souvent référence aux applications radars. En vision par ordinateur, on parle plutôt de "détection d'objets". L'idée générale derrière ces expressions est la détection de la présence d'une instance ou d'une classe d'objets, et le cas échéant de localiser le ou les éléments détectés. La littérature regorge de méthodes de détection sur des images d'intensité. On distingue deux catégories d'approche pour la détection : "par modèle d'apparence" (cf. Section 1.3.1) et "par le mouvement" (cf. Section 1.3.2). Ces deux types d'approche peuvent d'ailleurs être mises en œuvre simultanément de manière complémentaire. Plus en rapport avec notre problématique, les sections 1.3.3 et 1.3.4 font respectivement un bilan sur les méthodes de détection RGB-D et sur les méthodes de détection utilisées pour les insectes.

1.3.1 Détection par modèle d'apparence

Les approches de détection par modèle d'apparence font souvent appel à l'apprentissage supervisé. Celui-ci fonctionne classiquement en deux étapes :

- **Apprentissage "hors ligne"** : un jeu de données d'apprentissage est élaboré à partir d'exemples variés de l'objet à détecter (e.g. visages d'une personne sous différents angles). Ces données d'apprentissage servent à élaborer un modèle générique (e.g. classifieur) capable de représenter la classe de l'objet (e.g. visage de la personne). Le modèle peut aussi être raffiné à l'aide d'exemples négatifs (e.g. visage d'autres personnes).

- **Reconnaissance "en ligne"** : l'image est parcourue à la recherche de patches ressemblant au modèle appris lors de la phase hors ligne. Différentes mesures de similarités peuvent être utilisées.

Notons que le modèle doit être suffisamment générique pour englober les variations d'apparence au sein de la classe, mais pas trop pour ne pas retourner trop de faux positifs.

Ainsi, les approches vont des plus simples (e.g. correspondance de motifs par corrélation, approches plus générales de type TFAC) aux plus évoluées (e.g. sac de mots [Sivic 2005] s'appuyant sur des caractéristiques SURF [Bay 2006], SIFT [Lowe 1999], Kaze [Alcantarilla 2012], Harris [Harris 1988] pour la détection de points d'intérêts, par exemple combiné à un descripteur de type ZNCC¹). On note que la méthode par cascade de classifieurs faibles proposée par [Viola 2001] fait toujours figure de référence dans le domaine.

Ce qu'il faut retenir sur la détection par modèle d'apparence :

- **points forts** : construction automatique du modèle de classe par apprentissage ; fonctionne sur des images "hors séquence" ;
- **points faibles** : souvent peu de maîtrise directe du modèle ; sensibilité aux variations d'apparence.

1.3.2 Détection par le mouvement

Les approches de détection par le mouvement reposent sur le principe de soustraction d'arrière plan. L'idée est d'utiliser au sein d'une séquence vidéo, les images précédant et suivant l'image en cours pour modéliser l'arrière plan. Le modèle d'arrière plan intègre les éléments statiques de la scène, à l'opposé de l'avant plan qui contient les éléments en mouvement. L'avant plan est généralement calculé pour chaque image en soustrayant l'arrière plan. Les éléments en mouvement que comporte l'avant plan sont en fait des régions qui peuvent être isolées par extraction de composantes connexes.

Il existe de nombreuses approches de modélisation d'arrière plan, allant des plus basiques (différence d'images simple ou double [Kameda 1996], ou médiane adaptative [McFarlane 1995]), aux plus complexes (modèle Gaussien simple [Wren 1997] ou de mélange [Zivkovic 2006]).

Nous recommandons aux lecteurs de consulter [Sobral 2014] d'Andrews Sobral pour un état de l'art détaillé des méthodes de soustraction d'arrière plan ainsi que leur performance respective. Également, on note qu'un grand nombre d'implémentations sont disponibles au sein de la BGS Library².

Ce qu'il faut retenir sur la détection par le mouvement :

- **points forts** : pas de modèle défini d'apparence, ce qui évite les problèmes liés aux variations (e.g. forme, texture) ;
- **points faibles** : fonctionne uniquement sur des images issues d'une séquence et détecte, par nature, uniquement les objets en mouvement ; construction manuelle des modalités de détection (e.g. segmentation, seuils).

1. Les descripteurs fondés sur le ZNCC (pour Zero mean Normalized Cross-Correlation) sont utilisés pour faire de la corrélation de motifs. Ils sont simples et rapides à mettre en œuvre.

2. BGS Library disponible au téléchargement à l'adresse : <https://github.com/andrewssobral/bgslibrary>

1.3.3 Détection sur des données RGB-D

La littérature témoigne d'un grand nombre de travaux sur la détection de cibles à partir de séquences d'images 2D. En revanche, la démocratisation des dispositifs d'acquisition 3D (ou plutôt RGB-D) étant relativement récente, la littérature sur la détection de cibles en 3D pour des données "image d'intensité/carte de profondeur" est plutôt restreinte.

On commence à voir apparaître des descripteurs RGB-D pour la détection et la classification d'objets. Par exemple [Blum 2012] propose un descripteur appelé "Convolutional K-means". [Bo 2013], quant à lui, propose le HMP (pour Hierarchical Matching Pursuit).

Également, plus proche de nos aspirations, il existe quelques travaux utilisant la fusion d'information d'intensité et de profondeur, mais ces derniers sont plutôt destinés à détecter des cibles volumineuses (e.g. piétons [Howard 2007, Steiger 2009]) pour lesquelles l'information de profondeur est disponible à coup sûr. Par exemple [Steiger 2009] propose d'appliquer indépendamment sur l'image d'intensité et la carte de profondeur une méthode de soustraction d'arrière plan. Puis, à l'issue d'une fusion, un masque d'avant plan est calculé. Et finalement les cibles sont extraites par la recherche de contours (region growing) contrainte par les informations de profondeur.

Dans un autre domaine, [Goktuk 2005] utilise la carte de profondeur renvoyée par une caméra TOF pour faire de la classification d'objets, profitant ainsi de l'avantage de données indépendantes aux conditions d'éclairage, aux ombres ou à la texture. Toujours à partir des données issues de caméras temps de vol, [Swadzba 2008, Jiwon 2009] proposent quant à eux, des méthodes de détection d'obstacles pour la navigation. Aussi, [Gemeiner 2009] utilise le couple intensité/profondeur pour détecter les coins pertinents présents dans une scène.

1.3.4 Détecteurs utilisés pour les insectes (scènes en 2D)

Détection par modèle d'apparence : Dans [Miranda 2012], les auteurs mettent en œuvre la méthode de Viola-Jones avec un classifieur entraîné pour la détection d'abeilles. Dans [Chen 2012], des abeilles taguées sont détectées grâce à l'utilisation d'une transformée de Hough circulaire. Toujours pour la détection d'abeilles taguées, [Hendriks 2012] détecte la marque (point blanc) collée sur le dos des abeilles par une combinaison de laplaciens de gaussienne, suivie d'une érosion et d'un seuillage.

On note également un certain nombre de travaux de classification, allant donc bien au delà de la détection simple. Pour reconnaître des types de libellule, [Yuefang 2007] propose d'utiliser des invariants de Hu. Quant à [Huang 2009], il propose d'extraire des caractéristiques SURF sur des images d'insecte. Pour reconnaître des papillons, [Li 2009] propose d'appliquer une méthode de régression spectrale (plus performante qu'une Analyse en Composantes Principales dans ce contexte) pour réduire la dimension de l'espace de caractéristiques.

Dans un contexte non-déterministe, certaines méthodes de suivi de cibles, telles que le "Filtrage Particulaire" utilisent directement une mesure de vraisemblance entre le modèle et l'observation. C'est le cas par exemple de [Khan 2004a] qui utilise un modèle pré-calculé d'apparence d'abeille (appelé "eigenbee"), ou [Veeraraghavan 2008, Maitra 2009] qui utilisent des modèles adaptatifs d'apparence et pourvus de contraintes géométriques.

Détection par le mouvement : Les méthodes de détection d'insectes s'appuyant sur la soustraction d'arrière plan sont particulièrement efficaces lorsque les conditions (e.g. éclairage, couleurs, activités) sont contrôlées. Cela suffit par exemple pour détecter des four-

mis [Balch 2001] ou des abeilles [Kimura 2014] en conditions contrôlées de laboratoire. Toujours en conditions contrôlées, [Straw 2010, Liu 2012] détectent respectivement des drosophiles et des colibris par soustraction d'arrière plan effectuée simultanément sur plusieurs caméras pour ensuite fusionner les informations. Dans un autre registre, [Kimura 2011] propose une quantification vectorielle pour la détection d'abeilles individuelles dans une zone densément peuplée.

En revanche, en situation non contrôlée, une approche par soustraction d'arrière plan simple suffit rarement. Par exemple, pour détecter des abeilles en vol devant la ruche, [Campbell 2008] capture des séquences en niveaux de gris à l'aide d'une caméra classique positionnée à l'entrée de la ruche, et ajoute une étape de filtrage s'appuyant sur une contrainte de forme (i.e. ellipses multi-échelles). Pour détecter des chauves-souris en vol, [Theriault 2010] a recours à l'infrarouge via trois caméras filmant la scène simultanément, chacune des sources étant associée à une méthode de soustraction d'arrière plan adaptative suivi d'une étape de fusion de données. Dans [Reda 2013], la détection des fourmis est assurée par une soustraction d'arrière plan, mais les résultats étant trop limités, le suivi des cibles en aval est effectué de manière semi-automatique.

Au sein de la section 1.4 suivante, nous proposons une nouvelle méthode de détection par le mouvement à partir des données RGB-D.

1.4 Solution proposée : une segmentation hybride

À l'issue de l'état de l'art présenté dans la section 1.3, nous avons constaté un manque au niveau des méthodes de détection de petites objets à partir de séquences RGB-D. Notre objectif n'est pas la classification, mais la détection et la localisation de cibles en 3D, ainsi nous sortons du cadre des descripteurs RGB-D pour la classification présentés dans la section 1.3.3.

Dans cette section, nous proposons notre méthode de segmentation HIDS (pour Hybrid Intensity Depth Segmentation), adaptée aux informations d'intensité et de profondeur. On retrouve déjà ce concept de fusion dans [Steiger 2009]. Mais, contrairement à ces travaux, notre méthode s'appuie sur trois sources d'information :

- le mouvement, par différence d'images d'intensité successives ;
- le mouvement, par soustraction d'arrière plan d'intensité ;
- la position, par soustraction d'arrière plan de profondeur.

ce qui permet de détecter les cibles, en 3D lorsque c'est possible ou simplement 2D lorsque l'information de profondeur n'est pas exploitable. Cette approche offre également des possibilités intéressantes en termes de détection de cibles statiques (ou à faible mouvement).

1.4.1 Complémentarité des données d'intensité et de profondeur

Images d'intensité seules : Les méthodes de soustraction d'arrière plan, aussi performantes soient-elles, possèdent toutes leurs limites. Au vu des contraintes de notre application (cf. Section 1.2.1) et des tests préliminaires effectués, les valeurs d'intensité sont grandement affectées par des variations récurrentes d'éclairage. On observe également de nombreuses ombres portées ainsi que des réflexions visibles sur les ailes des abeilles par exemple. Les méthodes de soustraction d'arrière plan dites "adaptatives" sont difficilement paramétrables. En effet,

des seuils de détection de mouvement trop hauts auront tendance à favoriser l'intégration de bruit dans l'arrière plan. À l'inverse, des seuils trop bas auront tendance à surestimer le mouvement et ainsi augmenter le nombre de régions potentiellement assimilables à des cibles. Un compromis sur le seuil n'est pas toujours satisfaisant.

Cartes de profondeur seules : La carte de profondeur (ou carte des disparités) est calculée par un algorithme d'appariement de zones texturées entre les deux images d'intensité capturées par la caméra stéréovision. Ces cartes contiennent des "trous" (zones non appariées) dont les causes ont été énoncées précédemment dans la section 1.2.2.3.

Sur ces zones non appariées, il se cache peut-être des cibles. Dans des conditions optimales, une cible est représentée par un pic (zone en relief) sur la carte de profondeur. Mais dans des conditions difficiles, les cibles peuvent être présentes sur des régions non appariées. Néanmoins, une région non appariée ne correspond pas forcément à une cible. La figure 1.6 illustre différents effets observables sur la carte de profondeur. La majorité des zones non appariées sont dues à la différence de point de vue entre les deux objectifs, où l'un observe par exemple une cible en vol et l'autre observe ce qu'il y a derrière la cible du point de vue du premier objectif.

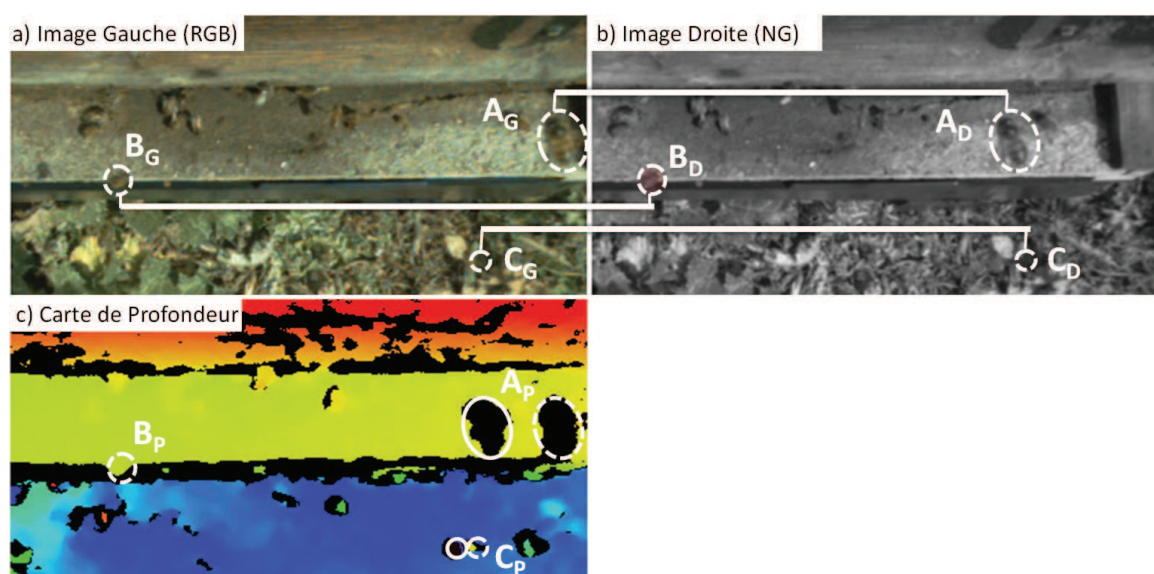


Figure 1.6 – Différents effets observés sur la carte de profondeur sur une scène capturée par la caméra stéréovision TYZX G3 Evo. a) image en couleur capturée par l'objectif de gauche, b) image en niveaux de gris capturée par l'objectif de droite, c) carte de profondeur calculée à partir de a) et b). L'abeille A (cible rapide) n'est pas correctement appariée entre a) et b) ce qui résulte en une première zone noire sur c). De plus, dû à sa hauteur par rapport à l'arrière plan (ici la planche d'envol) on observe une seconde zone noire sur sa gauche (marquée par l'ellipse en trait continu à gauche de A_P). Cet effet a été illustré à l'aide la figure 1.4. L'abeille B (cible statique) est correctement appariée, et sa proximité par rapport à l'arrière plan ne provoque pas de zone noire. L'abeille C (cible moyennement rapide), est largement devant l'arrière plan tout comme l'abeille A, elle est donc représentée par deux zones noirs sur c). Par contre, on voit que la zone noire de droite comporte tout de même de l'information de profondeur, ce qui témoigne d'un appariement partiellement correct.

Couple intensité/profondeur L'idée à l'origine de notre méthode de segmentation HIDS est de tirer parti des couples intensité/profondeur. L'intérêt étant de pouvoir détecter facilement les cibles dont la profondeur est disponible (pic sur la carte de profondeur), mais également de détecter les cibles dont la profondeur n'a pas pu être calculée lors de la phase d'appariement. La figure 1.7 illustre le principe utilisé. On recherche d'abord sur la carte de profondeur les zones non appariées (correspondant à des cibles potentielles), puis on confirme ensuite grâce aux images d'intensité (i.e. mouvement), si ces zones correspondent à des cibles. On note que sur l'image d'intensité, les changements soudains d'éclairage et les ombres portées sont inéluctablement considérés comme du mouvement, ce qui pourrait retourner de nombreux faux positifs. Cependant, il y a de grandes chances pour que les zones non appariées, dont la cause est le manque de texture, deviennent finalement appariables en présence de mouvement. En effet, le mouvement correspond à un changement de texture, ce qui offre des conditions optimales pour la phase d'appariement. C'est finalement cette idée de complémentarité permettant d'éviter les fausses détections qui fait la force de notre méthode de segmentation HIDS.

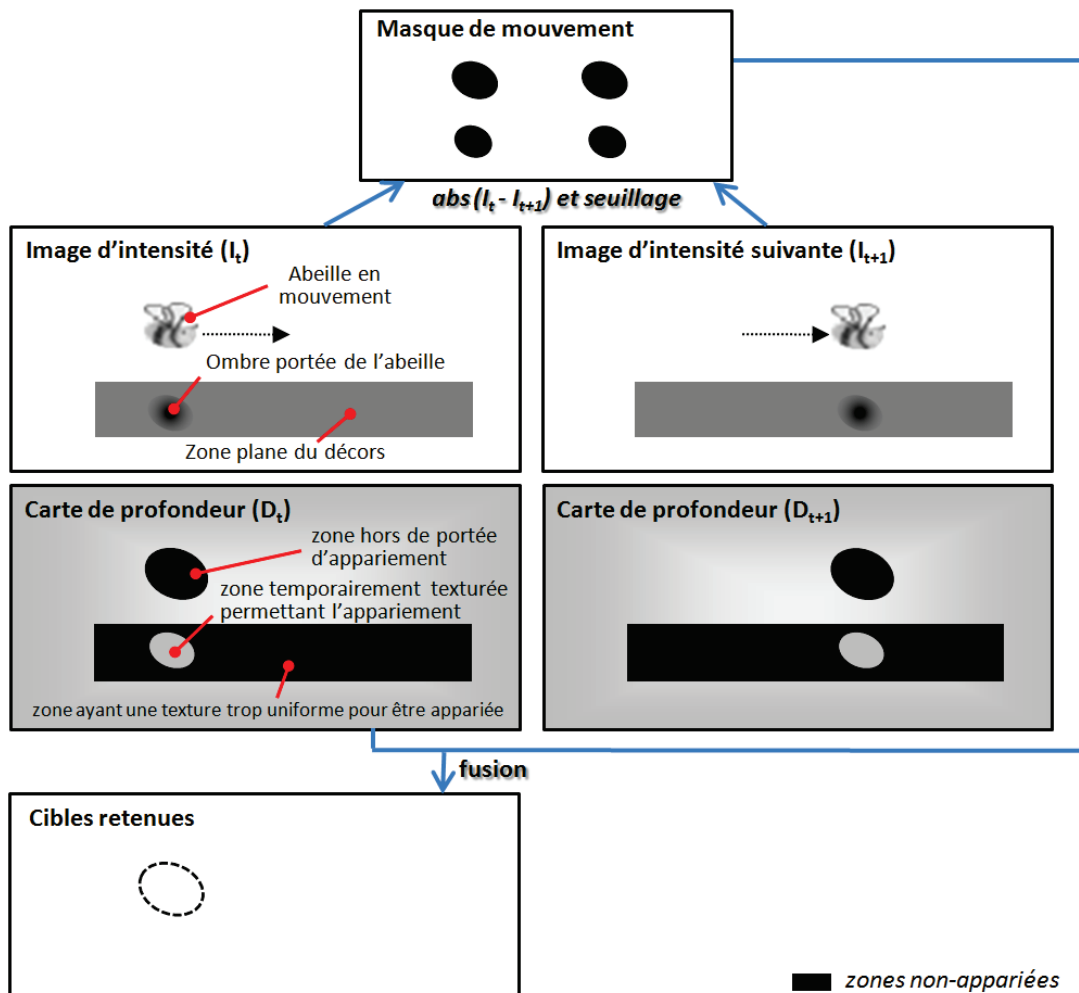


Figure 1.7 – Schéma illustrant la complémentarité intensité/profondeur pour la détection de cibles non appariées (e.g. zone hors de portée d'appariement).

1.4.2 Algorithme HIDS (Hybrid Intensity Depth Segmentation)

Cette section détaille notre méthode de segmentation HIDS de manière formelle. Notons que la méthode présentée peut être vue comme une extension des méthodes classiques reposant sur la modélisation adaptative de l'arrière plan. L'ajout étant l'utilisation de l'information de profondeur pour guider l'évolution du modèle de l'arrière plan.

La caméra stéréovision fournit une paire d'images d'intensité (gauche et droite) et une carte de profondeur (ou carte des disparités) associée. Soit, $I_{t,u,v}$ l'intensité du pixel au temps t à la position (u, v) , et $D_{t,u,v}$ sa distance par rapport à la caméra au temps t . Nous visons ci-dessous à calculer deux masques binaires à partir de I et D :

- le masque $DDTM$ (pour Determined Depth Target Mask) composé de plusieurs régions, chacune englobant une cible dont la profondeur peut être calculée ;
- le masque $UDTM$ (pour Undetermined Depth Target Mask) composé de plusieurs régions, chacune englobant une cible pour laquelle la profondeur ne peut pas être retrouvée.

1.4.2.1 Calcul de DDTM

Le masque $DDTM$ (Determined Depth Target Mask) est obtenu par une soustraction entre D et l'arrière plan de profondeur calculé dans DBG (pour Depth Background, voir ci-après) :

$$\forall_{t,u,v}, DDTM_{t,u,v} = \begin{cases} 1 & \text{if } D_{t,u,v} - DBG_{u,v} > \Delta d \\ 0 & \text{otherwise} \end{cases} \quad (1.2)$$

où Δd est un seuil. Une opération morphologique d'ouverture est ensuite appliquée pour supprimer le bruit présent sur la carte de profondeur D . Les zones de D ou DBG non définies (absence d'information sur la profondeur) ne sont pas prises en compte. Elles seront traitées à l'aide de UDM (voir ci-dessous).

L'arrière plan de profondeur DBG est calculé à partir de plusieurs images de profondeur via une approche médiane temporelle non-évolutive :

$$\forall_{u,v}, DBG_{u,v} = \text{median}\{D_{t_0,u,v}, D_{t_0+\Delta t,u,v}, \dots, D_{t_0+k\Delta t,u,v}\} \quad (1.3)$$

avec k étant le nombre d'images utilisées pour le calcul de la médiane, et Δt l'espacement (nombre d'images) à considérer entre deux images extraites à partir de la séquence. Contrairement aux valeurs d'intensité, les valeurs de profondeur sont globalement stables dans le temps et ne sont pas influencées par l'éclairage. Un léger effet de "scintillement" de l'ordre de quelques millimètres est engendré par les imperfections de l'algorithme d'appariement. Un simple moyennage dans le temps permet d'estimer la valeur de profondeur réelle avec une précision relativement bonne. La qualité de l'arrière plan de profondeur DBG dépend des conditions d'encombrement. L'utilisation d'un nombre d'images k élevé pour le calcul de la médiane et d'un Δt inter-image élevé permettent d'améliorer la robustesse de l'arrière plan de profondeur, et permet en particulier de faire abstraction des cibles dites "de passage", c'est à dire non stagnantes. La figure 1.8 illustre sur un exemple le calcul de $DDTM$.

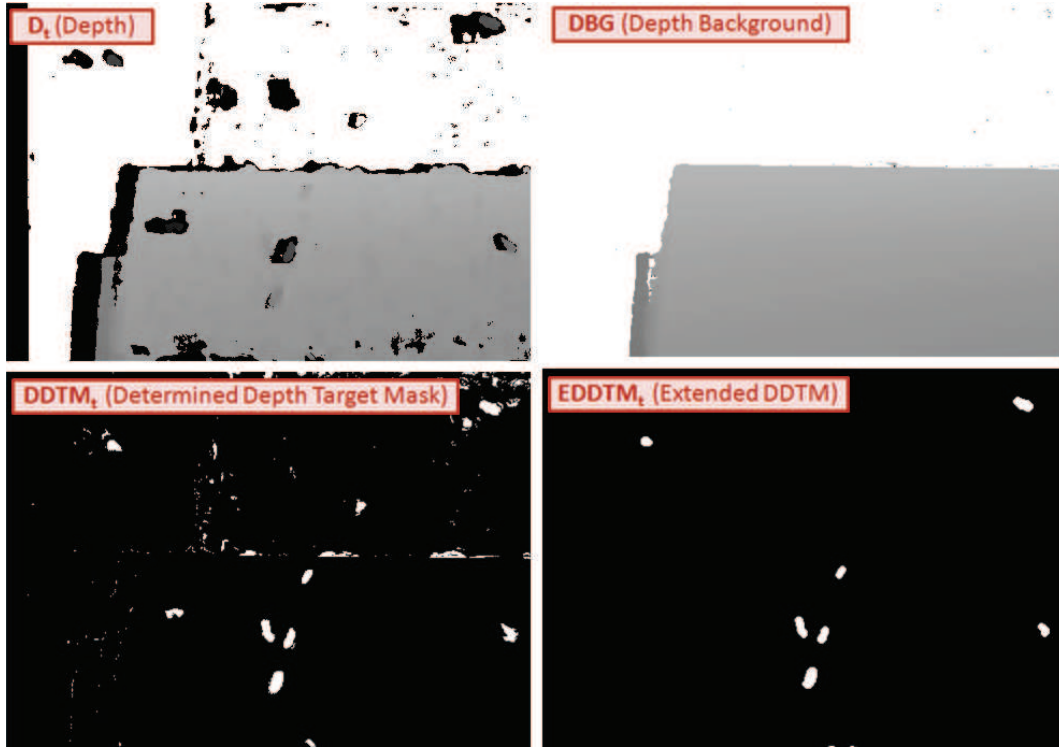


Figure 1.8 – Illustration du calcul de DBG et $DDTM$ à partir d'un exemple de séquence capturée à l'entrée d'une ruche. Les seuils utilisés sont $k = 100$ images, $\Delta t = 20$ images, $\Delta d = 1$ cm, et pour le calcul de $EDDTM$ (voir le calcul de FG), \circ est l'opération morphologique d'ouverture utilisant l'élément structurant S_4 de forme circulaire de 4 pixels de rayon.

1.4.2.2 Calcul de $UDTM$

Pour le calcul de $UDTM$, il est nécessaire de calculer $IAMM$ (pour Intensity Absolute Motion Mask) et UDM (pour Undetermined Depth Mask). Ce dernier contient les régions de la carte de profondeur D ayant une profondeur indéterminée mais exclut les régions de l'arrière plan de profondeur DBG ayant une profondeur indéterminée. UDM est calculé par l'intersection de D et DBG :

$$\forall_{t,u,v}, UDM_{t,u,v} = \begin{cases} 1 & \text{if } (D_{t,u,v} = e) \cap (DBG_{u,v} = e) \\ 0 & \text{otherwise} \end{cases} \quad (1.4)$$

où e est la valeur assignée par la caméra stéréovision aux pixels dont la profondeur n'est pas déterminée, et \cap l'opération de conjonction logique.

Ensuite, $IAMM$, correspondant aux régions de mouvement par rapport l'arrière plan d'intensité, est calculé à partir de la distance absolue entre IBG (voir ci-dessous) et I . Une opération morphologique de fermeture est appliquée afin d'élargir les régions potentielles de mouvement et de les fusionner avec leurs homologues proches. $IAMM$ est calculé par :

$$\forall_{t,u,v}, IAMM_{t,u,v} = \begin{cases} 1 & \text{if } ((|IBG_{t,u,v} - I_{t,u,v}| \oplus S_1 \ominus S_2) * M) > \Delta m \\ 0 & \text{otherwise} \end{cases} \quad (1.5)$$

où \oplus est l'opération morphologique de dilatation utilisant l'élément structurant S_1 , \ominus est l'opération morphologique d'érosion utilisant l'élément structurant S_2 , $*$ est l'opération de convolution avec le filtre moyennneur M , et Δm le seuil pour la binarisation. Notons que S_1 doit être plus grand que S_2 .

IBG (l'arrière plan d'intensité médian temporelle évolutif) utilisé pour le calcul $IAMM$ est initialisé par :

$$\forall_{u,v}, IBG_{t_0,u,v} = \text{median}\{I_{t_0,u,v}, I_{t_0+\Delta t,u,v}, \dots, I_{t_0+k\Delta t,u,v}\} \quad (1.6)$$

où tout comme pour le calcul de DBG , k est le nombre d'images utilisées pour le calcul de la médiane, et Δt est l'espacement (nombre d'images) à considérer entre deux images extraites à partir de la séquence.

La figure 1.9 illustre sur un exemple le calcul de $IAMM$.

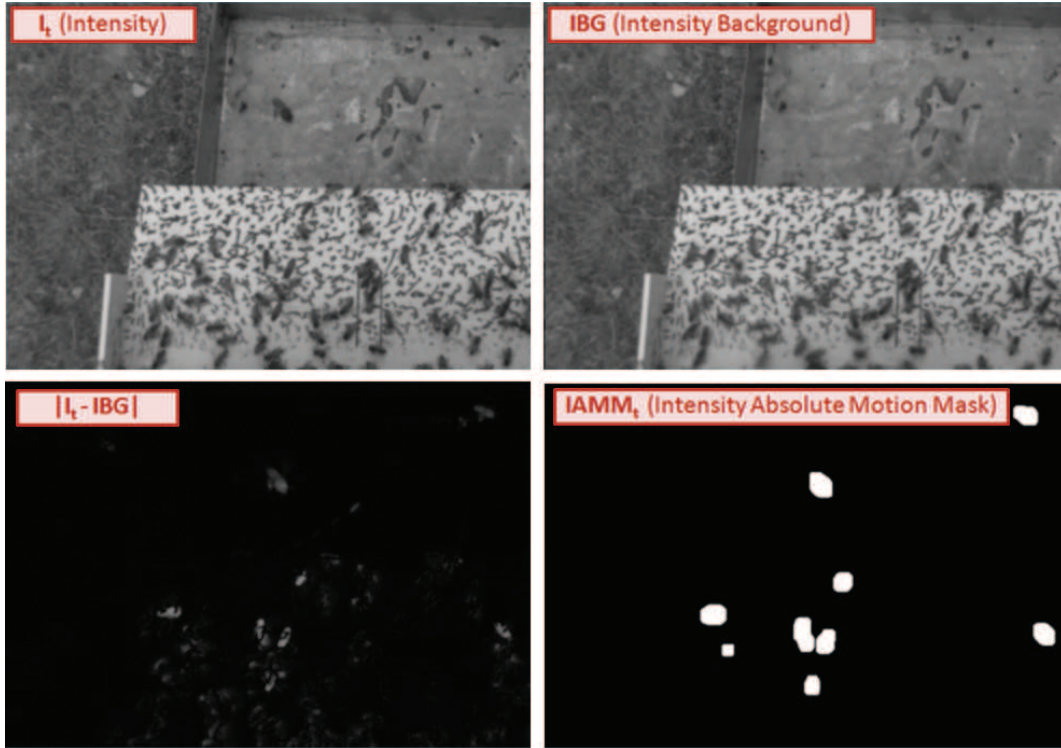


Figure 1.9 – Illustration du calcul de $IAMM$ à partir d'un exemple de séquence capturée à l'entrée d'une ruche. Les paramètres utilisés sont : l'élément structurant circulaire S_1 de rayon 10 pixels, l'élément structurant circulaire S_2 de rayon 8 pixels, le filtre moyennneur M de côté 6 pixels, et $\Delta m = 35$.

Finalement, $IAMM$ est calculé par l'intersection de UDM (auquel on retire $DDTM$) et $IAMM$:

$$\forall_{t,u,v}, UDTM_{t,u,v} = (\neg DDTM_{t,u,v} \cap UDM_{t,u,v}) \cap IAMM_{t,u,v} \quad (1.7)$$

La figure 1.11 illustre sur un exemple le calcul de $UDTM$.

1.4.2.3 Calcul de FG

Pour le calcul de FG (pour Foreground), il est nécessaire en plus de UDTM de calculer les masques IRMM (pour Intensity Relative Motion Mask) et EDDTM (pour Extended DDTM). Le masque d'intensité de mouvement relatif $IRMM$ comprend les régions en mouvement observées à partir d'un couple d'images d'intensité successives. Il est calculé par :

$$\forall_{t,u,v}, IRMM_{t,u,v} = \begin{cases} 1 & \text{if } (|I_{t,u,v} - I_{t-1,u,v}| \circ S_3) > \Delta rm \\ 0 & \text{otherwise} \end{cases} \quad (1.8)$$

où \circ est l'opération morphologique d'ouverture utilisant l'élément structurant S_3 pour élargir les zones potentielles de mouvement, et Δrm est un seuil de binarisation. La figure 1.10 illustre sur un exemple le calcul de $IRMM$.

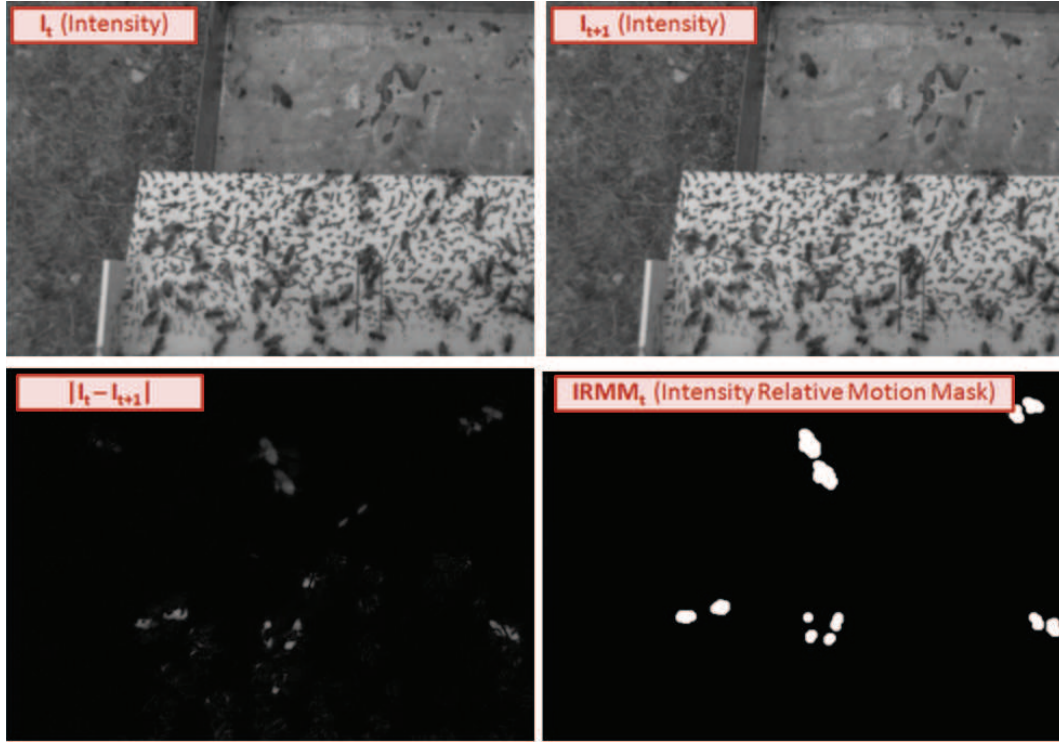


Figure 1.10 – Illustration du calcul de IRMM sur une séquence capturée à l'entrée d'une ruche. Les paramètres utilisés sont l'élément structurant circulaire S_3 de rayon 8 pixels, et le seuil $\Delta rm = 10$.

Le masque de premier plan FG inclut toutes les régions correspondant à des cibles potentielles, à l'exception des régions ne montrant aucun mouvement. FG est calculé par :

$$\forall_{t,u,v}, FG_{t,u,v} = EDDTM_{t,u,v} \cup (UDTM_{t,u,v} \cap \neg(UDTM_{t,u,v} \cap IRMM_{t,u,v})) \quad (1.9)$$

où $EDDTM$ est obtenu en appliquant une opération morphologique de dilatation sur $DDTM$, cela permet d'assurer que l'ensemble des cibles soit inclus dans FG .

$$\forall_{t,u,v}, EDDTM_{t,u,v} = DDTM_{t,u,v} \circ S_4 \quad (1.10)$$

La figure 1.11 illustre sur un exemple le calcul de FG . FG est utilisé pour effectuer une adaptation sélective de IBG étant définie par :

$$\forall_{t,u,v}, IBG_{t+1,u,v} = (\delta \cdot FG_{t,u,v}) \cdot I_{t,u,v} + ((1 - \delta) \cdot FG_{t,u,v}) \cdot IBG_{t,u,v} \quad (1.11)$$

où δ est un taux d'apprentissage utilisé pour l'adaptation.

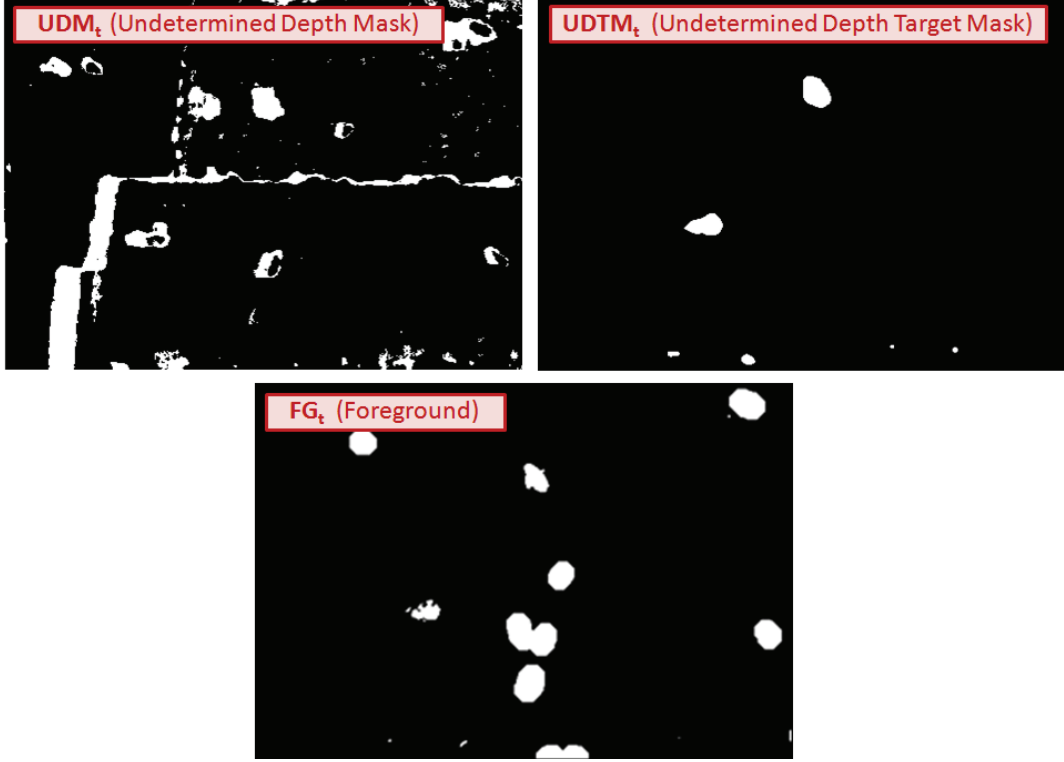


Figure 1.11 – Illustration du calcul de $UDTM$ et de FG sur une séquence capturée à l'entrée d'une ruche.

1.4.2.4 Synthèse sur la segmentation HIDS

La figure 1.12 illustre sous forme schématique l'algorithme de segmentation décrit ci-dessus. La figure 1.13 accompagne ce schéma par des exemples issus d'une séquence RGB-D capturée devant la ruche. La méthode de segmentation présentée est particulièrement adaptée pour des séquences capturées en extérieur en raison de la présence d'éléments naturellement texturés, ce qui offre des conditions optimales pour l'appariement stéréo. Les masques $DDTM$ et $UDTM$ sont utilisés dans la section suivante pour l'extraction des cibles en vol

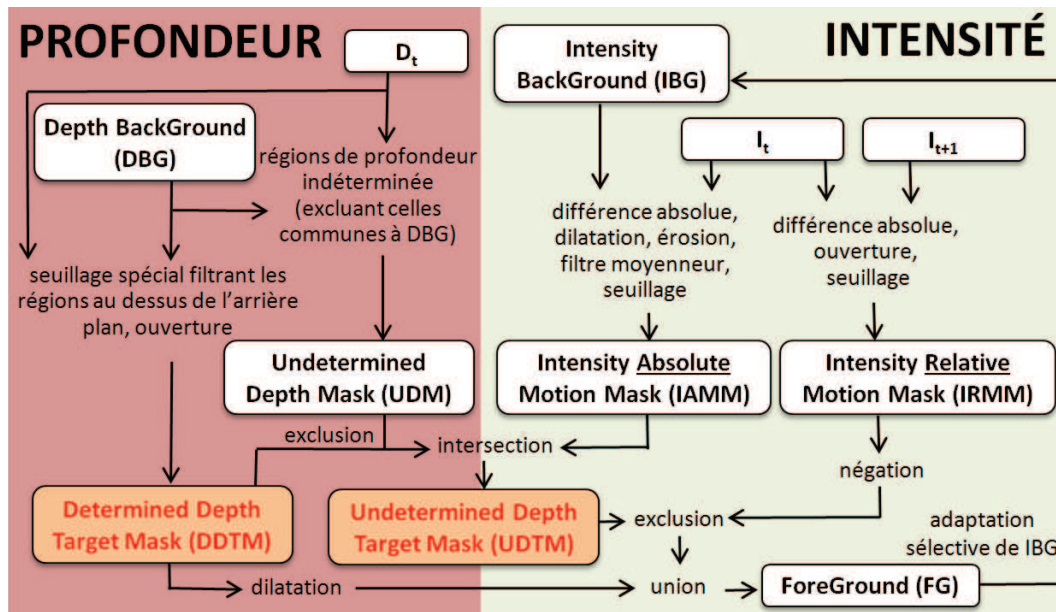


Figure 1.12 – Algorithme de la segmentation HIDS sous forme schématique.

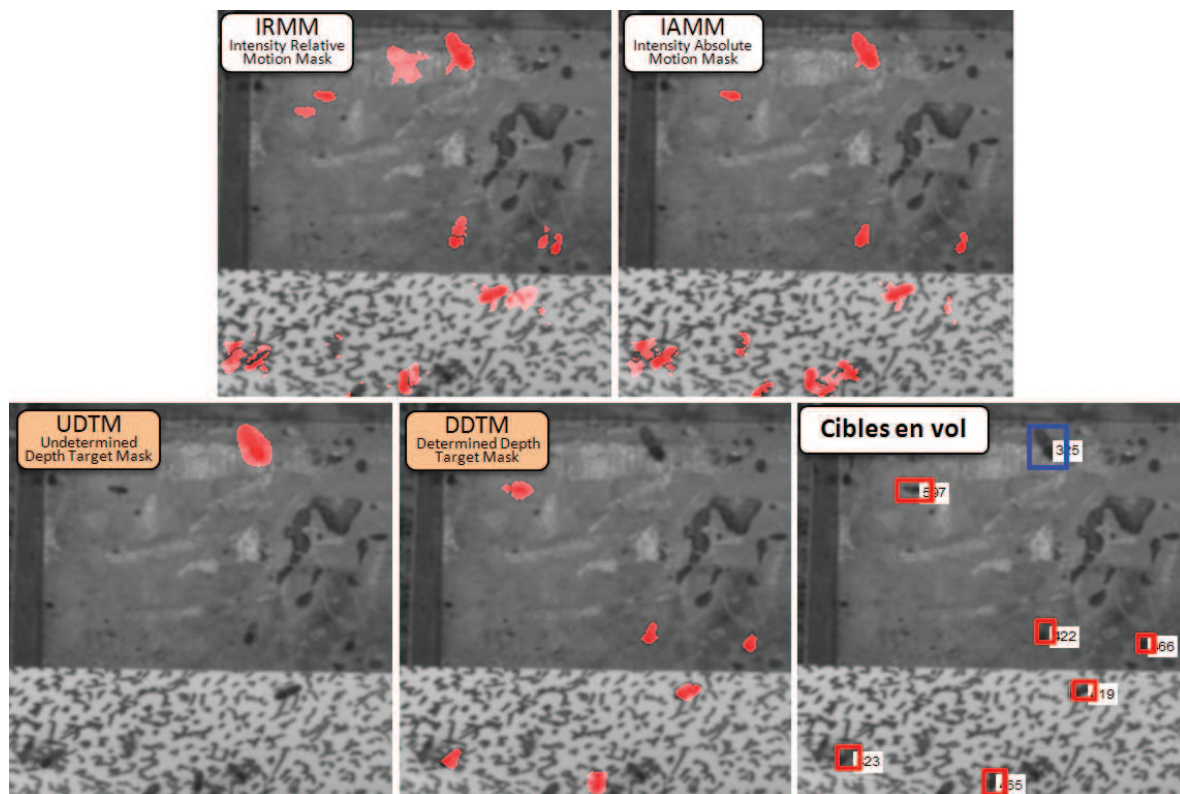


Figure 1.13 – Aperçu des étapes clés de l'algorithme HIDS sur une image issue d'une séquence RGB-D capturée devant la ruche. Les différents masques sont représentés par un filtre rouge transparent superposé sur l'image en niveaux de gris. L'étape "Cibles en vol" est effectuée en suivant la méthode d'extraction de cibles présentée dans la section suivante.

1.4.3 Extraction des cibles

Nous venons à l'aide de la section précédente de détailler l'algorithme permettant de calculer deux masques binaires *DDTM* et *UDTM*. Ces masques sont composés de régions distinctes assimilables à de potentielles cibles. *DDTM* représente les cibles dont l'information de profondeur peut être calculée, et à l'opposé, *UDTM* représente les cibles dont l'information de profondeur n'est pas disponible.

Les régions sont ensuite extraites par recherche de composantes connexes. Une cible est modélisée par le centroïde (ou barycentre) de la région. Pour les cibles issues de *DDTM*, la profondeur est calculée par la médiane des valeurs de D comprises dans la région concernée par la cible. De plus une ellipse englobante est estimée pour chaque région, informant ainsi sur la taille et l'orientation de la cible. La figure 1.14 montre des exemples d'ellipses extraites pour les cas suivants : profondeur disponible et indisponible, pour des cibles sur un arrière plan encombré ou non.

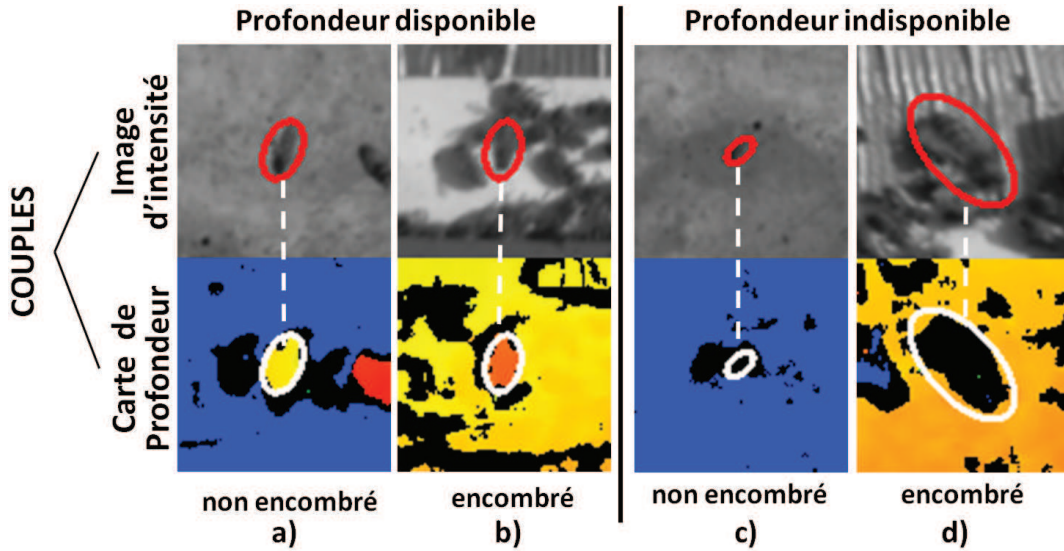


Figure 1.14 – Quatre exemples de cibles segmentées, avec les ellipses correspondantes sur les 4 couples d'images d'intensité/profondeur. La profondeur peut être estimée pour les cas a) et b), mais pas pour c) et d) où la cible est respectivement trop loin, et trop proche de la caméra (en dehors de portée d'appariement). Contrairement aux cas a) et c), les cas b) et d) correspondent à une zone encombrée induisant des difficultés pour modéliser l'arrière plan.

À l'issue de la segmentation HIDS, des statistiques sur la profondeur et la taille des ellipses ont été collectées. Nous avons identifié une relation exploitable entre la profondeur et la longueur de l'axe majeur de l'ellipse englobante. Cette relation apparaît linéaire à l'échelle de notre application, mais devrait théoriquement ne plus l'être sur une plage de profondeur plus grande. La figure 1.15 montre que, la relation peut être approximée par deux fonctions (μ et σ fonctions de la distance d , notées respectivement μ_d et σ_d par la suite). En se basant sur les données collectées à chaque unité de profondeur, nous avons estimé par régression polynomiale les coefficients de ces deux fonctions. De fait, plus la distance entre ce modèle et une observation est élevée, plus on peut émettre des doutes sur la véracité de l'observation. Ainsi, nous définissons le degré de vérité qu'une observation soit une fausse alarme par la

fonction d'appartenance suivante :

$$m_d^{FA}(s) = 1 - e^{-\frac{(s - \mu_d)^2}{2\sigma_d^2}} \quad (1.12)$$

où d est la profondeur et s la longueur de l'axe majeur de l'ellipse englobante.

À propos des cibles dont la profondeur n'est pas définie, notre idée initiale fut d'estimer la profondeur à partir de la longueur de l'axe majeur. Mais cette hypothèse a été invalidée étant donné l'écart-type trop important du modèle. Cette fonction d'appartenance sera finalement utilisée dans le chapitre suivant pour le suivi de cibles.

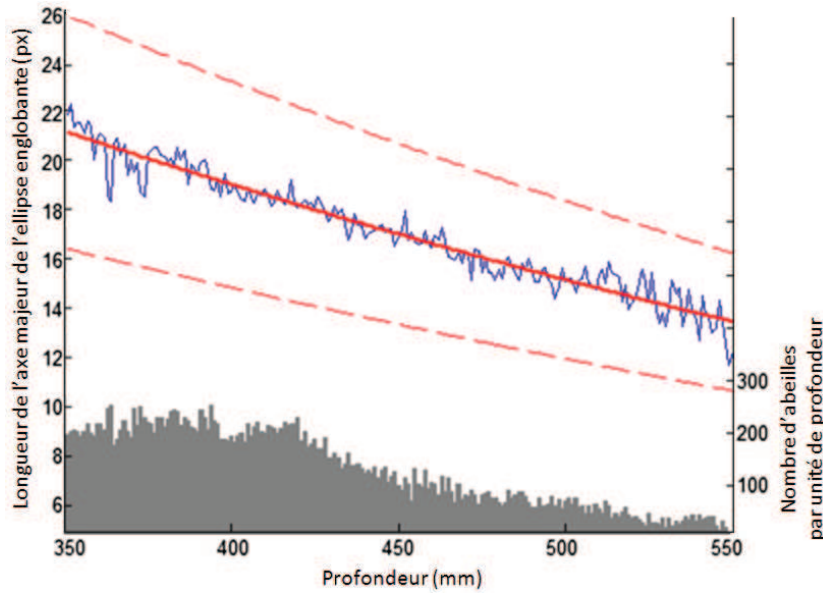


Figure 1.15 – Relation moyenne longueur/profondeur pour une cible de type "Abeille", calculée à partir de 50 000 exemples. La longueur étant celle de l'axe majeur de l'ellipse englobante. L'histogramme indique le nombre d'exemples utilisé par unité de profondeur. La courbe pleine (en bleu) représente la taille médiane en fonction de la profondeur. La courbe pleine et celles en pointillés (en rouge) représentent respectivement l'estimation du modèle longueur/profondeur, et l'écart-type associé.

1.5 Évaluation de la méthode de segmentation HIDS

Nous proposons dans cette section d'évaluer notre méthode de segmentation HIDS sur des séquences vidéos d'abeilles à l'entrée de la ruche. La figure 1.16 montre le dispositif d'acquisition stéréovision utilisé lors des premiers tests effectués à l'INRA Magneraud.

L'évaluation porte sur deux indicateurs : le taux de faux positifs (ou de fausses alarmes) et le taux de vrais négatifs (ou de détections manquées). Pour pouvoir calculer ces indicateurs, il est nécessaire de créer en amont une vérité terrain associée à notre jeu de données.



Figure 1.16 – Dispositif d'acquisition stéréovision fixé à l'entrée d'une ruche. Cette manipulation fut effectuée à l'INRA Magneraud dans le cadre de la première campagne de tests en 2012.

1.5.1 Élaboration de la vérité terrain

Pour chaque image, la vérité terrain doit répertorier la totalité des abeilles présentes dans le champ de la caméra. Malgré la pénibilité de la tâche, nous optons pour une procédure d'annotation manuelle.

1.5.1.1 Outil d'annotation

Pour l'occasion, nous avons développé un outil d'annotation en ligne (cf. Annexe A.1), à partir duquel les participants doivent dessiner des ellipses englobantes autour des abeilles, image après image. Ceci de façon à renseigner la position 2D ainsi que la longueur approximative des cibles. Pour se faire, l'utilisateur peut visionner l'image d'intensité, la carte de profondeur, et le mouvement par rapport au couple "intensité/profondeur" précédent. Ces trois sources d'information sont complémentaires et toutes nécessaires (à l'image de notre méthode HIDS). L'outil propose aux utilisateurs des couples d'intensité/profondeur extraits aléatoirement à partir de séquences vidéo mises à disposition. Le tirage aléatoire des couples d'images permet d'assurer une représentation exhaustive des situations.

1.5.1.2 Annotation en triple aveugle

Le processus d'annotation est long et laborieux mais n'est pour autant pas infallible. Les séquences vidéo sont capturées en conditions naturelles, et les différents jeux de lumière

donnent parfois lieu à des situations ambiguës du point de vue de l'observateur.

Pour limiter les biais liés aux erreurs d'appréciation humaines, nous avons mis en place un processus d'annotation en triple aveugle, où chaque image est annotée par trois individus différents. Les trois sources d'annotations sont prises en considération de la manière suivante. Une cible est considérée comme réelle si au moins deux des trois observateurs y font référence. On considère que deux annotations A_1 et A_2 font référence à une même observation O_1 si le centre d'une ellipse (correspondant par exemple à A_1) est inclus dans l'autre ellipse (correspondant à A_2) et vice versa. Ainsi, nous avons observé sur notre jeu de test que 4.8% des ellipses annotées (45 sur 928 ellipses) sont considérées comme des erreurs d'appréciation. Cette approche nécessite trois fois plus de travail qu'une approche classique, mais elle permet néanmoins d'assurer la qualité de la vérité terrain.

1.5.1.3 Données annotées

Notre vérité terrain d'abeilles en vol devant la ruche a été créée à partir de 500 images issues de neuf séquences de 15 minutes chacune. Les séquences ont été capturées sous différentes conditions (cf. Figure 1.17). Également, on peut apercevoir sur la figure 1.18, l'extrême variabilité des abeilles annotées au niveau de la forme, de l'apparence et de la taille.

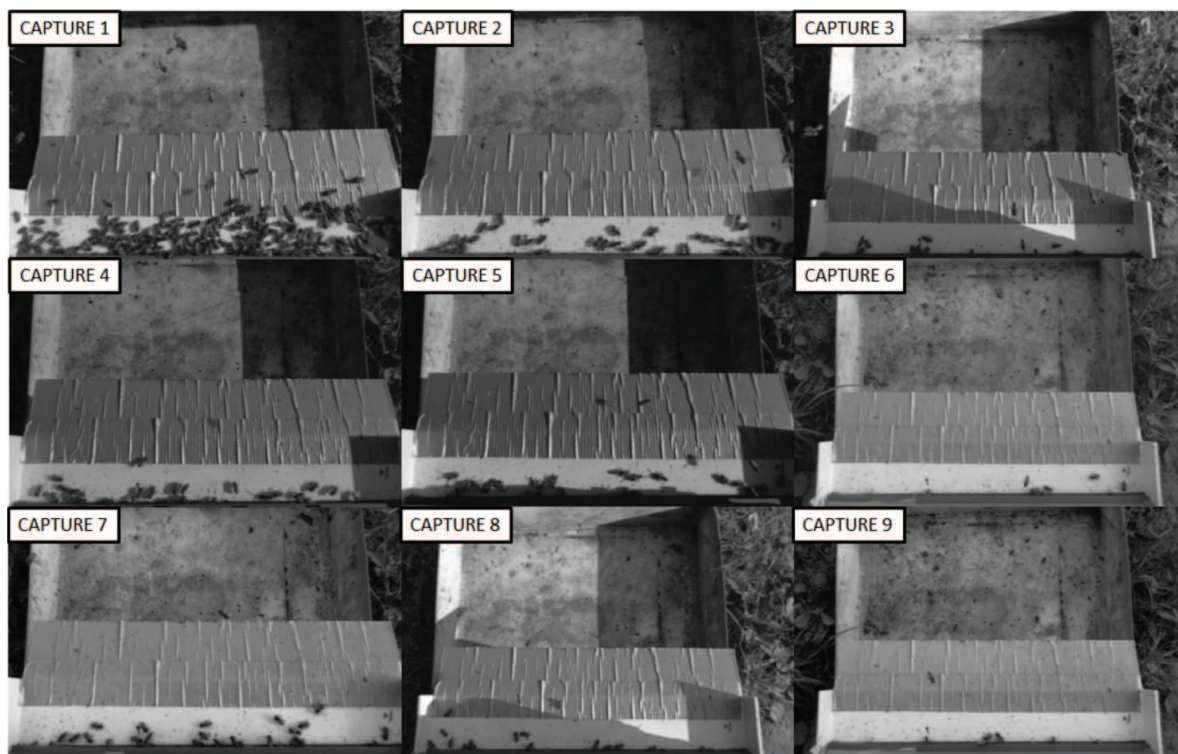


Figure 1.17 – Images extraites des neuf séquences utilisées pour la construction de la vérité terrain. Ces captures représentent une large variété de conditions (activité, nervosité des abeilles, éclairage, réflexions, ombres portées).



Figure 1.18 – Échantillons de 100 abeilles annotées à partir des neuf captures disponibles (images triées par tailles).

1.5.2 Résultats

1.5.2.1 Segmentation HIDS

Les résultats présentés dans cette section sont issus de comparaisons entre les données de segmentation HIDS (automatique) avec les données de la vérité terrain (annotations manuelles). Le tableau 1.4 détaille les résultats de notre segmentation HIDS dans les conditions des neuf captures mentionnées dans la section précédente. Premièrement, on note qu'un éclairage instable a tendance à faire augmenter le taux de fausses alarmes. Ensuite, on constate un taux de cibles non détectées plus élevé lorsque l'activité en vol est importante. Finalement les taux moyens de détections manquées et de fausses alarmes sont respectivement de 4.15% et 19.54% (moyennes pondérées par le nombre d'échantillons issu de chaque capture).

L'intérêt du caractère hybride de notre segmentation HIDS est encore une fois appuyé dans les sections 1.5.2.4 et 1.5.2.3 au sein desquelles nous proposons des tests de détection via des approches fondées sur l'intensité ou la profondeur de manière indépendante.

Capture	Activité	Éclairage	Ombres	Caméra	VN	FP
Capture 2	++	TI	+++	L	11.02%	28.81%
Capture 5	+	I	++++	L	0.00%	27.97%
Capture 7	+	I	+++	L	0.00%	17.80%
Capture 8	++	S	+++	H	6.78%	16.10%
Capture 1	++	S	+++	L	1.69%	15.25%
Capture 4	++	S	+++	L	0.00%	9.32%
Capture 3	+++	S	++	H	8.47%	5.93%
Capture 6	++	TS	+	H	0.85%	3.39%
Capture 9	++	TS	+	H	0.00%	0.00%
Moyennes pondérées					4.15%	19.54%

Tableau 1.4 – Résultats de la segmentation HIDS par capture (conditions variables). Activité ou nombre moyen d'abeilles en vol par image (+ moins de 1, ++ entre 1 et 8, +++ plus de 8), Éclairage (TI = Très Instable, I = Instable, S = Stable, TS = Très Stable), Ombres (intensité des ombres), Caméra (L = Position basse, H = 10 cm plus haut), VN (Taux de Vrais Négatifs, ou de cibles non détectées), FP (Taux de Faux Positifs, ou de détections de cibles inexistantes).

1.5.2.2 Répartition spatiales des erreurs

La figure 1.19 montre que la zone de la planche d'envol est particulièrement touchée par les fausses alarmes. Cela est dû aux mouvements ininterrompus des abeilles sur la planche, ou aux ombres portées des abeilles en vol. Ces zones de mouvement sont détectées sur les images d'intensité et sont alors candidates pour être associées à d'éventuelles zones sans profondeur recouvrable, ce qui produit finalement des fausses alarmes (cf. 1.4).

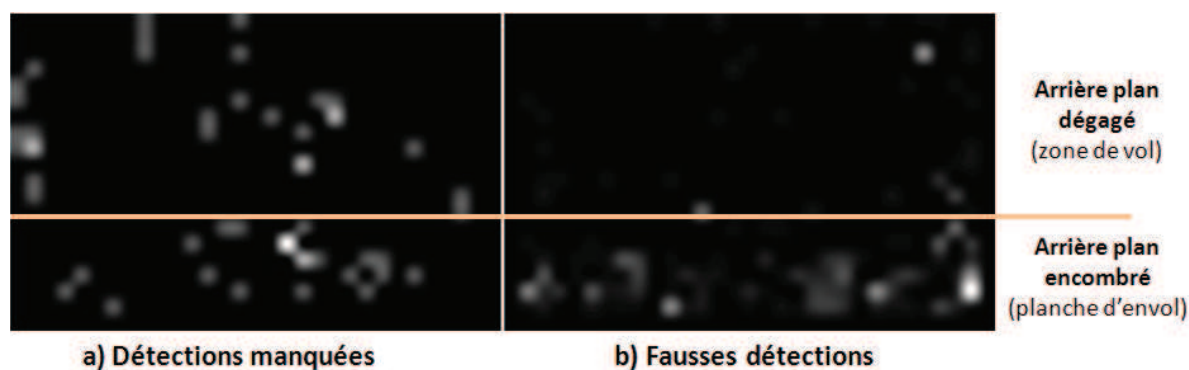


Figure 1.19 – Représentation de la distribution spatiale des détections manquées (a) et des fausses détections (b). Les détections manquées sont uniformément réparties dans l'espace tandis que les fausses détections ont plutôt tendance à se produire au niveau de la planche d'envol.

1.5.2.3 Tests à partir de l'information de profondeur seulement

Notre méthode de segmentation HIDS offre des résultats de l'ordre de 4.15% de détections manquées et 19.54% de fausses alarmes (cf. Tableau 1.4). On note que 90% des fausses alarmes sont issues de régions dépourvues d'information de profondeur, et donc causées par

une surestimation du mouvement présent dans les images d'intensité.

Ainsi, nous avons comparé les résultats moyens de notre segmentation HIDS avec une seconde approche de segmentation s'appuyant uniquement sur les informations de profondeur. Cette seconde approche permet seulement de détecter les cibles dont la région correspondante a été correctement appariée lors du calcul de la carte de profondeur. Ainsi, toutes les cibles non appariées ne sont donc pas détectées, soit près de 30% des cibles dans notre application. Cette seconde approche renvoie des résultats de 48.09% de détections manquées et 3.78% de fausses alarmes, soit un taux de fausses alarmes plus bas, en dépit d'un taux de détections manquées sensiblement plus élevé. Ceci témoigne bien de l'importance d'utiliser l'information d'intensité pour la détection des abeilles en vol.

Finalement, dans le cadre de notre segmentation HIDS, les seuils optimaux utilisés pour la détection du mouvement entraînent une hausse du taux de fausses alarmes, mais c'est le prix à payer pour pouvoir détecter 95% des abeilles annotées dans la vérité terrain.

1.5.2.4 Tests à partir de l'information d'intensité seulement

Détection par le mouvement Nous avons tout de même effectué des tests de détection de cibles par une méthode adaptative de soustraction d'arrière plan. Cela fonctionne correctement dans un seul cas bien particulier : planche dégagée et aucune ombre portée, ce qui est trop contraignant dans le cadre de notre application.

Détection par modèle d'apparence Nous avons abordé dans l'état de l'art une piste concernant la détection de cibles via des méthodes s'appuyant sur l'apparence de type Viola-Jones [Viola 2001]. Nous avons effectué des tests en entraînant une cascade de classifieurs (via l'implémentation OpenCV). Pour cela nous avons fourni en guise d'exemples positifs un échantillon de 300 imagerie d'abeilles issues de la vérité terrain. Et en guise d'exemples négatifs, nous avons fourni un ensemble d'images d'arrière plan sans abeille. Les résultats furent très pauvres, nous avons donc rapidement abandonné cette piste. Ceci s'explique par l'extrême hétérogénéité des imagerie d'abeilles (cf. Figure 1.18). De plus, la résolution moyenne des imagerie est probablement insuffisante.

Nous avons aussi tenté de détecter les abeilles en vol et sur la planche à l'aide d'une LPE [Meyer 1994] (Ligne de Partage des Eaux) appliquée sur le gradient de niveau de gris, mais sans résultat probant.

1.5.3 Synthèse sur l'influence des paramètres

Cette section propose, via le tableau 1.5, une synthèse de l'influence des différents paramètres utilisés dans notre méthode de segmentation HIDS. Par exemple, les ailes transparentes des abeilles produisant des artéfacts qu'il est nécessaire de traiter avec des opérations morphologiques adaptées. De manière générale, les paramètres liés à la taille des éléments structurants sont sensibles (à 1 ou 2 unités près relativement à la taille des cibles), tandis que ceux liés à la robustesse de la médiane ou au seuil de profondeur sont moins sensibles.

Param./Val. défaut	Influence : valeurs inférieures (−) et supérieures (+)
Seuil de profondeur : $\Delta d = 1$ cm	+ : meilleure robustesse vis-à-vis des imprécisions de la carte de profondeur, mais masque DDTM potentiellement sous-estimé. − : risque de surestimation du masque DDTM.
Robustesse des médianes : - échantillon : $\Delta t = 20$ img - quantité : $k = 100$ img	Le rapport $\Delta t/k$ reste stable pour couvrir la même longueur Δt + : plus rapide, mais moins robuste. Δt − : prise en compte de conditions plus variées, plus lent.
Seuils de binarisation : - Mvt. relatif : $\Delta rm = 10$ - Mvt. absolu : $\Delta m = 35$	Influences liées aux éléments structurants ci-dessous + : sous-estimation du mouvement (masques IRMM & IAMM) − : surestimation du mouvement (masques IRMM & IAMM)
Tailles éléments struct. : - dilatation : S_1 de 10 px - érosion : S_2 de 8 px - convolution : M de 6 px - ouverture : S_3 de 8 px - ouverture : S_4 de 4 px	Ces opérations travaillent dans le domaine spatial de manière absolue. La taille des éléments structurants est donc définie en fonction de la résolution de l'image, de la taille et de la nature des cibles à détecter. Des valeurs trop importantes tendent à considérer les cibles comme du bruit. Et inversement, des valeurs trop faibles tendent à augmenter le taux de fausses alarmes.
Taux d'apprentissage : $\delta = 0.2$	+ : Cibles peu mobiles potentiellement plus rapidement incrustées dans le modèle de d'arrière plan. - : Modèle d'arrière plan (IBG) difficilement mis à jour (e.g. éclairage, salissures).

Tableau 1.5 – Synthèse des paramètres (et valeurs par défaut) utilisés dans notre méthode de segmentation HIDS.

1.6 Conclusion

Dans ce premier chapitre, nous avons posé les bases d'un système de vidéosurveillance capable de détecter des cibles rapides et de petite taille, évoluant dans un espace à trois dimensions. Le domaine applicatif présente de nombreux verrous scientifiques et techniques, simultanément présents, qu'il a fallu relever. Bien que la littérature soit déjà très riche, les contraintes applicatives ont nécessité l'apport de nouvelles contributions.

Du côté matériel, nous avons opté pour une caméra stéréovision (modèle TYZX G3 Evo³) dont nous avons détaillé la configuration et également discuté de ses avantages et inconvénients par rapport à d'autres technologies d'acquisition 3D.

Concernant le traitement d'images pour la détection de cibles, nous avons proposé une méthode générique de segmentation hybride intensité/profondeur que nous avons nommée segmentation HIDS. Cette méthode s'appuie sur la fusion d'information de mouvement (1 source s'appuyant sur une différence d'images successives à court terme, et une source définie par une soustraction d'arrière plan à moyen terme), ainsi que des informations de profondeur (une source obtenue par une soustraction d'arrière plan).

En guise de validation, nous avons testé l'ensemble de notre dispositif sur des séquences vidéo d'abeilles à l'entrée de la ruche. Les résultats sont satisfaisants, et offrent une base solide pour envisager un suivi des cibles de façon individuelle pour la collecte de trajectoires en 3D. Ceci fait l'objet du chapitre suivant.

3. La société TYZX a été rachetée par Intel Corporation en 2012. Le dispositif G3 Evo n'est donc plus en vente à l'heure actuelle.

Suivi 3D multi-cibles

Sommaire

2.1	Introduction	46
2.2	État de l'art des approches probabilistes	47
2.2.1	Contexte théorique	47
2.2.2	Formalisation du problème	48
2.2.3	Méthodes classiques de filtrage	49
2.2.4	Variantes existantes, pour quels contextes ?	52
2.2.5	Association de cibles pour le suivi par filtrage de Kalman	56
2.3	Solutions théoriques proposées	58
2.3.1	Filtrage de Kalman 3D et mesures 2D/3D	59
2.3.2	Suivi contraint par un modèle de scène	62
2.3.3	Suivi contraint par une base de données de trajectoires	65
2.4	Préparation de la mise en œuvre pour le suivi d'abeilles	68
2.4.1	Présentation de l'application	68
2.4.2	Configuration du filtre de Kalman	71
2.4.3	Modélisation de la planche d'envol	71
2.4.4	Construction d'une base de données de trajectoires d'abeilles en vol	73
2.5	Méthode d'évaluation et résultats	74
2.5.1	Mise en place d'une vérité terrain	74
2.5.2	Définition des modèles de scénario	75
2.5.3	Métrique pour l'évaluation	75
2.5.4	Résultats sur différents modèles de scénario	76
2.6	Conclusion	82

2.1 Introduction

Dans le chapitre précédent, nous avons présenté une méthode nommée "segmentation HIDS" capable de détecter des cibles en 3D à partir de séquences RGB-D. Le point fort de cette approche consiste en sa capacité à détecter les cibles même lorsque l'information de profondeur n'est pas fiable voire absente. Le cas échéant, les cibles sont localisées simplement en 2D.

En tant que deuxième maillon de la chaîne, ce chapitre propose d'aborder le problème du suivi de cibles, notamment à partir du mélange d'observations 2D/3D renvoyé par notre méthode de segmentation. Le calcul des trajectoires consiste principalement à filtrer le bruit inhérent aux observations et à associer ensemble les observations issues de mêmes cibles tout au long de leur existence au sein de la scène.

Problématique Nous posons ainsi la problématique suivante : *"comment à partir d'observations 3D, parfois incomplètes et bruitées, peut-on suivre simultanément dans l'espace 3D de multiples cibles volantes au mouvement d'apparence chaotique ?"*. Notons que la caractérisation du mouvement (stable ou erratique) est relative à l'échelle, à la résolution et à la fréquence d'observation (cf. Figure 2.1).

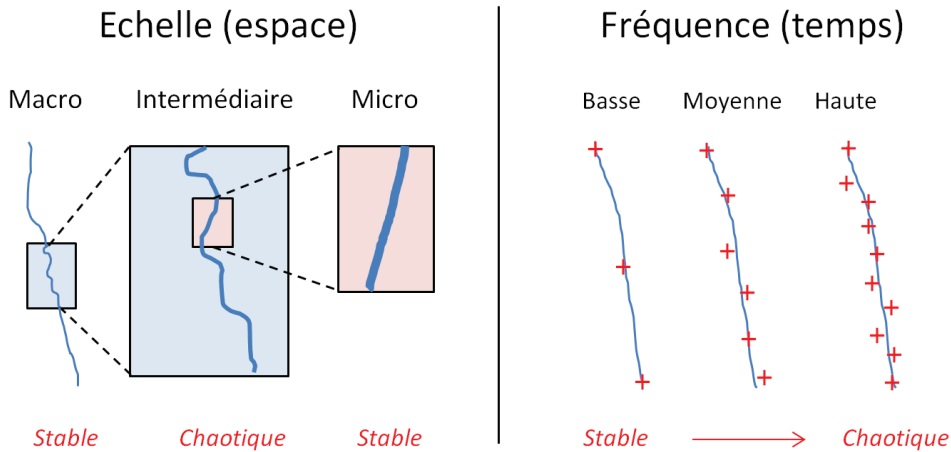


Figure 2.1 – Caractérisation ambiguë du mouvement au niveau spatio-temporel. Une même trajectoire peut finalement être qualifiée de : 1) stable au niveau "micro" grâce des opérations de filtrage malgré une fréquence d'observation élevée, 2) plutôt chaotique au niveau "intermédiaire", ce qui entraîne des difficultés en termes d'association "traces/observations", et 3) éventuellement stable au niveau "macro" où l'on peut retrouver une dynamique globale contextualisée par rapport à la scène.

Littérature La littérature comporte un nombre important d'études traitant des problèmes de suivi de cibles. On y trouve un large panel de solutions théoriques répondant à des problèmes divers et variés. En voici une liste non-exhaustive :

- suivi de cibles en n-dimensions (2 ou 3D) ;
- dans l'espace d'observation (e.g. repère image) ou bien dans l'espace réel ;
- avec gestion de plusieurs cibles simultanées ou non ;

- en connaissant le nombre de cibles à suivre ou non ;
- avec un nombre de cibles à suivre stable ou non ;
- où l'on connaît la dynamique des cibles ou non ;
- où l'on peut modéliser plus ou moins de précisément la dynamique des cibles (e.g. linéaire ou non) ;
- où le suivi s'appuie sur des observations éventuellement bruitées ;
- avec gestion possible des occultations ;
- avec prise en compte ou non de l'apparence des cibles (e.g. modèles adaptatifs).

Solution proposée En revanche, en ce qui concerne le suivi de cibles à partir d'observations mixtes 2D/3D, ou bien la prise en compte d'une dynamique de cible contextualisée par rapport à la scène, les solutions théoriques clefs en main se font rares. C'est à ces problèmes spécifiques que nous nous attaquons dans ce chapitre. Ainsi, dans le but d'y apporter une réponse, et par la même occasion de lever les verrous scientifiques associés, nous proposons d'apporter les trois contributions suivantes :

1. une méthode de projection permettant au filtre de Kalman défini en 3D de traiter des observations 2D temporairement définies dans un espace d'observation restreint (i.e. suppression d'une dimension) ;
2. deux méthodes distinctes s'appuyant sur des approches fondamentalement opposées, pour la prise en compte de l'influence de la scène sur la dynamique des cibles :
 - 2.a. méthode avec un fort *a priori*, s'exécutant rapidement : approche définie à partir d'une relation hypothétique entre la dynamique d'une cible et sa proximité avec un élément (ou zone d'intérêt) de la scène ;
 - 2.b. méthode avec peu d'*a priori*, demandant plus de ressources de calcul : approche s'appuyant sur un ensemble de trajectoires (i.e. historique représentant un ensemble de dynamiques possibles en tout point de la scène).

2.2 État de l'art des approches probabilistes

2.2.1 Contexte théorique

Le suivi d'objets (ou tracking) à partir de vidéos est un vaste domaine faisant état de très nombreux travaux et ceci depuis plusieurs décennies par les communautés "statistiques", "traitement du signal et automatique" et "vision par ordinateur". Les applications sont multiples, allant par exemple de l'interaction homme-machine à la réalité augmentée en passant par la surveillance, l'imagerie médicale ou l'édition vidéo. Naturellement, il n'existe pas de solution générique pour le suivi de cibles qui satisfasse l'ensemble des besoins existants. Ainsi, pour chaque application, il convient donc de trouver la méthode la mieux adaptée au problème. La figure 2.2 illustre les principales approches de suivi de cibles proposées dans la littérature. Le lecteur intéressé peut consulter [Yilmaz 2006] pour plus de détails sur les différentes approches.

Comme il a été précisé dans le chapitre précédent (cf. Section 1.2.1), nous cherchons à suivre des cibles d'une nature particulièrement difficile (i.e. de petite taille et ayant une dynamique erratique, similaires mais d'apparence instable dans le temps). Par élimination, l'approche de suivi la plus adaptée à notre problème s'avère être celle dite "probabiliste". Nous avons d'ailleurs dans ce contexte développé la méthode de segmentation HIDS relativement

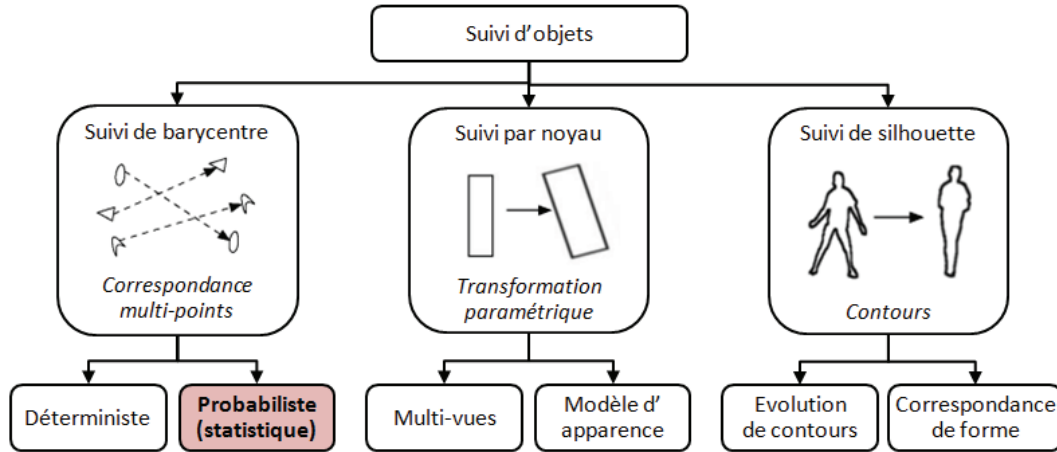


Figure 2.2 – Différentes approches de suivi de cibles. Notre étude se situe dans la catégorie "Probabiliste". Illustration adaptée de [Yilmaz 2006].

fiable et performante de détection de cibles (cf. Chapitre 1) dans le but de maximiser les chances de réussite de ce type d'approche. La nature des observations renvoyées par notre méthode de détection de cible nous placent finalement dans un contexte d'étude proche de celui des signaux radars.

2.2.2 Formalisation du problème

Les approches probabilistes sont fortement liées à la théorie de l'estimation. Le filtrage Bayésien (aussi appelé estimation récursive Bayésienne) est généralement utilisé pour estimer une trajectoire à partir d'une série d'observations. De manière formelle, le problème du suivi d'une cible se résume, pour chaque itération k , à trouver une estimation de la densité de probabilité *a posteriori* de l'état x_k (e.g. position 3D) d'une cible à partir d'un ensemble de mesures observées $Z_{1:k} = Z_1, \dots, Z_k$. La règle de Bayes adaptée au problème s'écrit :

$$P(x|Z_{1:k}) = \frac{P(Z_{1:k}|x)P(x)}{P(Z_{1:k})} \quad (2.1)$$

où $P(x|Z_{1:k})$ est la densité de probabilité *a posteriori* que l'on cherche à estimer, $P(Z_{1:k}|x)$ est la fonction de vraisemblance et $P(x)$ est la densité *a priori*. Une solution optimale à ce problème d'estimation peut être obtenue par un processus récursif en deux étapes :

- **la prédiction** permet à l'itération k une approximation *a priori* de l'état x de la cible sous la forme de la densité $P(x_k|Z_{1:k-1})$ conditionnellement à toutes les observations précédentes. Le calcul de cette densité s'appuie sur la densité *a posteriori* mise à jour $P(x_{k-1}|Z_{1:k-1})$ à l'itération précédente via l'étape de correction, ainsi que sur le modèle de dynamique de la cible observée via la densité $P(x_k|x_{k-1})$ issue de la fonction f_k de transition d'état. D'où l'équation :

$$P(x_k|z_{1:k-1}) = \int P(x_k|x_{k-1})P(x_{k-1}|z_{1:k-1})dx_{k-1} \quad (2.2)$$

- **la correction** permet de mettre à jour la densité conditionnelle prédite *a priori* $P(x_k|Z_{1:k-1})$

en prenant en compte la nouvelle observation Z_k , et la densité $P(z_k|x_k)$ (vraisemblance) issue de la fonction d'observation h_k . Soit l'équation de la densité *a posteriori* :

$$p(x_k|z_{1:k}) = \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{p(z_k|z_{1:k-1})} \quad (2.3)$$

avec la constante de normalisation, suivant le caractère markovien :

$$p(z_k|z_{1:k-1}) = \int p(z_k|x_k)p(x_k|z_{1:k-1})dx_k \quad (2.4)$$

La présence d'intégrales rend complexe le calcul de la solution présentée ci-dessus pour sa mise en pratique. La partie suivante présente les trois méthodes classiquement employées.

2.2.3 Méthodes classiques de filtrage

Aujourd'hui encore, le problème énoncé précédemment fait toujours l'objet de recherches dans le but de trouver une solution générique. Historiquement, les trois grandes approches suivantes ont retenu l'attention des chercheurs :

- **Filtre de Kalman** : depuis les années 60, le filtre de Kalman [Kalman 1960] est considéré comme la meilleure solution pour traiter les problèmes linéaires à bruit gaussien (une sous classe du problème original). Pendant ces 50 dernières années, et encore aujourd'hui, on voit apparaître de nouvelles méthodes s'appuyant sur ce principe. Celles-ci permettent notamment d'étendre le champ d'applications de la méthode originale.
- **Filtrage Particulaire** : à partir des années 90, on note un engouement des chercheurs pour les méthodes à particules [Gordon 1993], permettant notamment de traiter des problèmes non-linéaires. La puissance grandissante des ordinateurs a grandement contribué à la démocratisation des filtres à particules, et les applications sont aujourd'hui nombreuses.
- **Filtre PHD** : le filtre PHD (pour Probability Hypothesis Density) [Mahler 2003] fut initialement introduit par Mahler au début des années 2000 et fait toujours l'objet de nombreuses recherches. Le Filtre PHD propose d'intégrer dans un même algorithme les divers composants nécessaires au pistage (e.g. filtrage et gestion multi-cibles) que l'on a l'habitude d'utiliser de manière modulaire dans les approches classiques.

Les théories des approches classiques de la littérature (i.e. Filtrage de Kalman et Filtrage particulaires) sont brièvement abordées dans les deux sections suivantes. De plus amples détails sur leurs particularités et leurs extensions sont données plus loin dans la section 2.2.4. Nous n'abordons cependant pas la théorie du filtre PHD dans ce manuscrit. Nous avons préféré concentrer nos efforts sur des approches déjà bien documentées et maîtrisées, les rendant relativement plus simples à implémenter et à mettre en œuvre dans le cadre d'une chaîne de traitement comme la nôtre.

2.2.3.1 Filtrage de Kalman

Le filtre de Kalman classique [Kalman 1960] est une solution optimale conçue pour estimer l'état de systèmes linéaires gaussiens. Notons que dans le domaine de la vidéosurveillance, la définition du filtre au sens strict est applicable à une minorité d'applications de suivi de cibles. Néanmoins, le filtre de Kalman gère des systèmes perturbés par un bruit blanc gaussien au

niveau du modèle d'évolution d'état et au niveau du modèle d'observation, ce qui permet bien souvent son emploi dans des situations sortant de son cadre théorique optimal. Formellement, si l'on considère les bruits comme étant de nature stationnaire dans le temps, on pose les deux modèles suivants :

$$\begin{aligned} \text{Évolution d'état : } X_k &= AX_{k-1} + v \quad \text{avec } v = \mathcal{N}(0, R) \\ \text{Observation : } Z_k &= CX_k + \omega \quad \text{avec } \omega = \mathcal{N}(0, Q) \end{aligned} \quad (2.5)$$

où pour le modèle d'évolution d'état, X_k est le vecteur d'état à l'itération k , A la matrice de transition d'état, et v un bruit gaussien de moyenne nulle de covariance R . Pour le modèle d'observation, Z_k est le vecteur de mesure, C est la matrice de liaison avec le vecteur d'état X , et ω un bruit gaussien de moyenne nulle de covariance Q .

De manière récursive, le filtre de Kalman permet alors d'estimer l'état X_k d'un système dont la transition d'état est connue, ainsi que l'incertitude P_k associée à cette estimation. Ceci en prenant en entrée une série de mesures $Z_{1:k} = (Z_1, \dots, Z_k)$ observées ainsi que les bruits d'évolution d'état et de mesure définis en amont. La figure 2.3 illustre le mécanisme du filtre de Kalman. Cette approche est dite "detect before track" (ou "tracking-by-detection") puisqu'une phase de détection de cibles (ou segmentation) préalable est nécessaire afin de pouvoir fournir un vecteur de mesure en entrée du filtre.

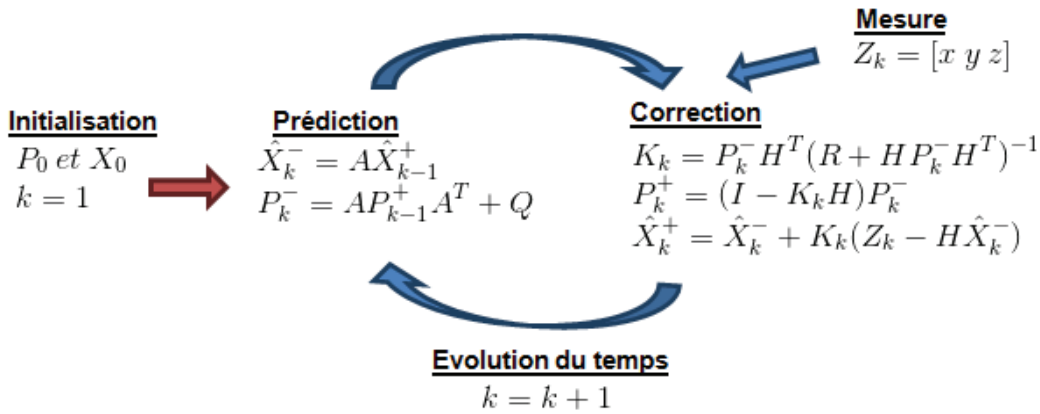


Figure 2.3 – Mécanisme récursif (Prédiction/Correction) du filtre de Kalman

Si l'état initial n'est pas connu de manière précise, on peut compenser en initialisant la matrice de covariance P_0 (ou d'incertitude) avec des valeurs relativement grandes, de cette façon le filtre va converger vers une bonne approximation de l'état réel. La vitesse de convergence du filtre dépend des modèles de bruit définis. Lorsque l'on connaît la spécification du système observé, les matrices de covariance de bruits R et Q peuvent être calculées analytiquement. Sinon, il est d'usage de les définir de manière empirique. Nous le verrons par la suite, dans le cas où les équations de modélisation du problème ne sont pas linéaires, il est possible d'estimer les paramètres du système à l'aide d'un filtre de Kalman étendu qui linéarise localement le problème.

2.2.3.2 Filtrage particulaire

Utilisées dans de nombreuses applications de suivi de cibles, les approches à base de particules [Doucet 2001, Arulampalam 2004] sont particulièrement adaptées aux problèmes hautement non-linéaires. L'idée générale du filtrage particulaire (FP aussi appelé méthode séquentielle de Monte Carlo, ou méthode par condensation) est de représenter la fonction de densité de probabilité *a posteriori* par un ensemble d'échantillons pondérés et tirés aléatoirement appelés "particules". Ainsi, suivant la théorie des grands nombres, plus le nombre de particules est élevé, plus l'estimation se rapproche de la solution optimale. De nombreuses variantes fondées sur cette approche existent. On note l'émergence de l'algorithme générique SIR (Sampling Importance Resampling) qui est aujourd'hui classiquement utilisé dans les applications de suivi de cibles. L'algorithme SIR [Gordon 1993] propose une méthode d'estimation récursive de l'état par particule. On note la densité $p(x_k|z_{1:k})$ *a posteriori* conditionnelle aux observations :

$$p(x_k|z_{1:k}) = \sum_{i=1}^N \omega_k^{(i)} \delta(x_k - x_k^{(i)}) \quad \text{avec} \quad \sum_{i=1}^N \omega_k^{(i)} = 1 \quad (2.6)$$

avec $x^{(i)}$ faisant référence à une particule de poids $\omega^{(i)}$. À chaque itération, la répartition des particules est effectuée à l'aide d'une fonction q dite d'importance, permettant de faire un échantillonnage préférentiel. Intuitivement, cela permet de répartir de manière uniforme les particules afin de représenter au mieux $p(x)$ tout en prenant en compte les poids des particules, avec la contrainte de garder l'ensemble des poids normalisés. La mise à jour des poids des particules est influencée par la vraisemblance de l'observation, $p(z_k|x_k^i)$. Formellement, les étapes suivantes sont effectuées, décrites par les équations :

$$\begin{aligned} \text{Échantillonnage indépendant :} \quad & x_k^{(i)} \sim q(x_k|x_{k-1}^{(i)}, z_k) \quad , \quad \forall k \\ \text{Mise à jour des poids :} \quad & \omega_k^{(i)} \propto \omega_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, z_k)} \quad , \quad \forall k \end{aligned} \quad (2.7)$$

L'état des particules évolue au cours du temps en fonction de la dynamique décrite par $p(x_k|x_{k-1})$. Finalement, on peut estimer l'état x_k par la moyenne pondérée des vecteurs d'état des particules, donnée par l'équation :

$$x_k = \sum_{i=1}^N \omega_k^{(i)} x_k^{(i)} \quad , \quad \forall k \quad (2.8)$$

Sans une dernière étape de ré-échantillonnage (assimilable à un filtre SIS), le filtre SIR aurait tendance à dégénérer [Kong 1994]. La dégénérescence se traduit par une majorité de particules associées à des poids nuls, provoquant ainsi un déséquilibre. Le filtre SIR propose, si nécessaire, de ré-échantillonner les particules à la fin de chaque itération, permettant ainsi de conserver un nombre significatif de particules efficaces.

En conclusion, le filtre particulaire explore par simulation un ensemble de solutions, et s'adapte au fur et à mesure aux observations. On considère cette approche comme étant de type "track before detect" ou (tracking-before-detection).

Et finalement, en général, on utilise le filtrage particulaire lorsque l'on est sûr que le Filtre Kalman est inefficace, ou lorsque l'on a un besoin précis (e.g. cibles déformables de type "blob"). On note également que malgré les définitions génériques du FP, il convient d'adapter l'algorithme en fonction de l'application.

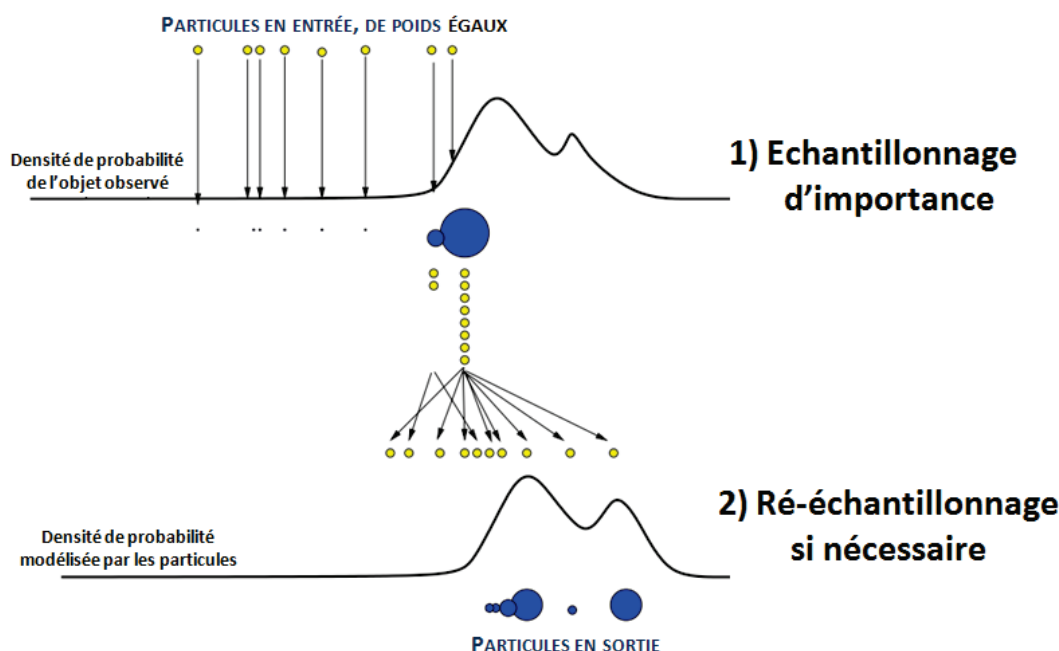


Figure 2.4 – Illustration du filtre SIR (Sampling Importance Resampling). Inspirée de la présentation "N.J. Gordon : Lake Louise : October 2003 - p. 23/47"

2.2.4 Variantes existantes, pour quels contextes ?

La section précédente a permis de poser les définitions originales des méthodes de suivi de cibles couramment utilisées, à savoir le filtre de Kalman et le filtrage particulaire. A savoir qu'il existe de nombreuses variantes et extensions à ces méthodes parmi lesquelles le choix n'est pas toujours évident. Nous proposons dans cette section de classer ces différents travaux (liste non-exhaustive) en s'appuyant sur des caractéristiques clés. Afin de garder un aspect pratique, chaque caractéristique est accompagnée d'un exemple applicatif.

2.2.4.1 Dynamique du système

- **Système linéaire** : Suivi de la position d'un avion ;
- **Système non-linéaire** : Suivi de la position d'un pendule.

Dans le cas où les fonctions de transition d'état f_k et d'observation h_k sont linéaires, le filtre de Kalman classique [Kalman 1960] offre une solution optimale. En revanche, dans le cas non-linéaire, le filtre de Kalman classique n'est plus adapté car il n'est plus possible de représenter les équations de transition sous forme matricielle. Le filtre de Kalman Etendu (EKF) [Blackman 1999] propose, à l'aide de matrices Jacobiennes, de linéariser les fonctions de transition localement à l'échelle du temps observé. De cette manière, il devient alors possible d'appliquer les équations du filtre de Kalman classique dans le cas non-linéaire.

Pour les cas hautement non-linéaires, le Unscented Kalman Filter (UKF) [Julier 1997] peut s'avérer plus performant. À la différence de EKF, UKF propage l'état à l'itération suivante via une transformation non-linéaire (Unscented Transform) sur $2n+1$ échantillons (appelés points "Sigma") où n est la dimension du vecteur d'état. Les points "Sigma" représentent la moyenne

et la covariance de la distribution gaussienne représentative de l'état. Cette approche permet ainsi d'éviter les erreurs de linéarisation induites par les calculs de Jacobiennes dans EKF. Il en résulte également un gain au niveau du temps de calcul par rapport à son homologue.

Les filtres particuliers [Doucet 2001, Arulampalam 2004] offrent une solution intéressante aux problèmes de suivi de cibles non linéaires. On obtient d'ailleurs, avec un nombre élevé de particules, une estimation plus précise que celle obtenue par EKF. Concernant les situations extrêmes (e.g. cibles ayant un mouvement brownien), le filtrage particulaire est capable de ressortir de bons résultats, là où les méthodes de type filtre de Kalman atteignent leurs limites.

Notons que les méthodes de suivi de cibles abordées ci-dessus nécessitent de définir un modèle de transition d'état. Il est aisé de modéliser la dynamique d'objets ayant une inertie importante (e.g. véhicules, avions), en revanche cela se complique lorsque l'on travaille avec de petites cibles capables d'accélération et de changements brusques de direction (e.g. humains, animaux, insectes). Parfois, nous ne pouvons pas définir un modèle de dynamique exacte. Dans la pratique, on se contente bien souvent d'appliquer des modèles de type "vitesse constante" ou "accélération constante", quitte à ce que le modèle ne soit pas en parfaite adéquation avec la dynamique réelle des cibles. Plusieurs approches permettent de palier ce problème. Par exemple, l'observation de la scène à plus haute fréquence [Chiron 2013a] permet de limiter l'aspect "brownien" de la dynamique des cibles.

2.2.4.2 Représentation multimodale

- **Mono-modale** : Suivi d'un poisson dans une rivière ;
- **Multi-modale** : Suivi d'un banc de poisson dans une rivière, le banc étant considéré comme une cible capable de se diviser temporairement.

Les approches de type Kalman se contentent d'estimer l'état d'une cible, par une approximation gaussienne de la distribution *a posteriori* (représentée par la moyenne X et la matrice de covariance P). Cette estimation *a posteriori* s'effectue lors de l'étape de correction, après la prise en compte par le filtre d'une nouvelle mesure. Des méthodes d'association de données telles que JPDA (pour Joint Probabilistic Data Association) [Bar-Shalom 1975] permettent de prendre en compte simultanément plusieurs mesures pour l'estimation d'une trace. Néanmoins, cette approche fusionne les différentes hypothèses d'association de la trace aux différentes observations pour ne finalement retourner qu'une approximation de distribution *a posteriori* sous forme d'une gaussienne.

En se rapprochant des concepts particuliers, on retrouve le filtre EnKF [Evensen 2003] pour (Ensemble Kalman Filter), une variante du Filtre de Kalman où la matrice de covariance est remplacée par des échantillons issues de la distribution gaussienne. Dans ce contexte, il existe des extensions s'appuyant sur des modèles de mélange de gaussiennes [Dovera 2011, Smith 2007] permettant d'étendre l'EnKF à des situations multimodales. Finalement, les approches particulières permettent par leur nature de représenter des distributions ayant plusieurs modes. Attention tout de même aux éventuelles dégénérescences (i.e. poids concentré sur une seule particule) qui peuvent diminuer le caractère multimodal du filtre. Plusieurs méthodes ont été proposées pour éviter ce phénomène. Parmi celles-ci, on note la linéarisation locale [Doucet 2001], le lissage par noyaux [Liu 2001], les FP auxiliaires [Pitt 1999].

2.2.4.3 Gestion de plusieurs hypothèses

- **Mono-hypothèse** : Suivi d'insectes au décollage (sur une vidéo) ;

- **Multi-hypothèse** : Suivi d'insectes au décollage (sur une vidéo), avec des ombres portées sur la surface d'envol.

Concernant les approches de type Kalman, il est possible de considérer plusieurs hypothèses grâce à des méthodes d'association de données de type MHT (Multi Hypothesis Tracker, cf. Section 2.2.5.2), originellement connues sous le nom d'algorithme de Reid's [Reid 1979]. L'idée est de conserver en mémoire dans un arbre d'hypothèses (à l'échelle de quelques itérations passées) un nombre fini des meilleures hypothèses d'association d'une "traces/observations". Ainsi les décisions logiques d'association de cibles sont effectuées de manière différée.

Contrairement aux approches de type filtre de Kalman, les approches particulières sont par nature multimodales, et chaque mode peut correspondre à une hypothèse. Des hypothèses peuvent alors apparaître, disparaître ou être maintenues en fonction des observations. Ce phénomène dépend grandement de la nature de l'observation, de la fonction de calcul de poids des particules et de la méthode d'échantillonnage des particules. Par exemple, si pour une cible à suivre, la caractéristique observée est sa couleur, des particules peuvent s'étendre sur n'importe quelles régions de couleur similaire, et ainsi former autant d'hypothèses que de régions en question. En revanche, si la caractéristique observée comprend également la position, les régions lointaines, bien que similaires ne seront plus considérées, ainsi le nombre d'hypothèse sera restreinte.

2.2.4.4 Contraintes externes

- **Sans contrainte externe** : Suivi d'une mouche en vol ;
- **Avec contraintes externes** : Suivi d'une voiture dans une ville (contrainte de rester sur la voirie).

Malgré des approches s'appuyant sur des modèles à hypothèse forte (e.g. *a priori* sur la dynamique), on constate généralement un fonctionnement complètement non contraint vis-à-vis de l'environnement (e.g. topologie de la scène). Pourtant, il est rare que l'état du système visite en pratique l'ensemble des valeurs offertes par la liberté du modèle. Concernant les applications de suivi de trajectoires, des propositions ont été faites pour prendre en compte les murs (e.g. suivi de piétons), ou les routes (e.g. suivi de convois). Cela peut se traduire par l'injection d'*a priori* supplémentaire sur l'environnement, conduisant à une adaptation contrainte des états prédits *a priori* et *a posteriori* [Simon 2002].

Concernant le filtrage particulière, il est plus simple de prendre en compte les contraintes externes, notamment en associant un poids nul aux particules dont l'état viole les contraintes de la scène [Arulampalam 2004] (e.g. position impossible, ou vitesse trop élevée). Dans certains travaux traitant de problèmes multi-cibles, on propose de prendre en compte des contraintes d'interaction entre les cibles au niveau des particules [Khan 2004b].

2.2.4.5 Approches multi-modèles

- **Mono-modèle** : Suivi de voitures en vitesse de croisière sur l'autoroute ;
- **Multi-modèle** : Suivi de cibles manœuvrantes.

Les approches multi-modèles proposent d'utiliser pour une même cible, plusieurs filtres simultanément, chacun correspondant à un modèle de suivi spécifique. Par exemple, ces ap-

proches sont très utilisées dans le cas de cibles manœuvrantes, que l'on a l'habitude de suivre à l'aide des trois modèles suivants :

- à l'arrêt,
- en croisière : modèle NCV (pour Nearly Constant Velocity),
- en manœuvre : modèle NCT (pour Nearly Coordinated Turn).

En sortie, les estimations des différents filtres sont combinées selon leur probabilité associées. On peut aussi dans certains cas choisir de ne garder que l'estimation du filtre qui semble la meilleure. L'implémentation "IMM" pour (Interactive Multiple Model) donnée par [Blom 1988] fait figure de référence dans le domaine. Toute la difficulté est de trouver les critères (e.g. probabilités) de transitions entre les états. Il existe des méthodes adaptatives (AIMM) qui permettent l'estimation de ces paramètres.

Les modèles SLDS (pour Switching Linear Dynamic System) permettent de décrire la dynamique d'un processus physique non-linéaire complexe, cela en changeant de modèle au cours du temps parmi un ensemble de modèles de dynamiques linéaires. La transition entre les modèles est définie par une matrice de transition Markovienne. On note l'utilisation de l'approche SLDS pour le suivi de cibles dans [Veeraraghavan 2008, Oh 2008].

Les approches multi-modèles permettent d'adapter le modèle de transition d'état en fonction de la situation. Notamment, en référence à la section "Contraintes externes" précédente, cela peut être utilisé pour basculer entre plusieurs modèles afin d'adapter la dynamique du système aux contraintes physiques de ce dernier et/ou aux contraintes environnementales de la scène. Cela peut aussi bien s'appliquer aux approches de type filtre de Kalman [Blom 1988] qu'aux approches particulières [Oudjane 1999, Boers 2003].

2.2.4.6 Gestion de plusieurs cibles

- **Mono-cible** : Suivi d'un poisson dans un aquarium ;
- **Multi-cible** : Suivi d'abeilles en vol devant leur ruche.

Dans les applications de vidéosurveillance, le cas optimal mettant en scène une seule cible visible du début à la fin est plutôt rare. On rencontre fréquemment des séquences présentant plusieurs cibles susceptibles de se croiser, d'être cachées par un élément de la scène, ou même d'entrer et de sortir du champ de la caméra. Dans le cadre d'une approche de type Kalman, on emploie classiquement des méthodes d'association de cibles de type plus proche voisin (NN pour Nearest Neighbor), ou plus évoluées, éventuellement la méthode GNN [Blackman 1986] (pour Global Nearest Neighbors). Dans le cas multi-hypothèses, nous l'avons vu précédemment, il est possible d'employer des méthodes de type MHT.

Il est possible de gérer le suivi de plusieurs cibles avec le filtrage particulière. Classiquement, on associe un filtre indépendant à chaque cible. Lorsque les cibles ont une apparence similaire, il devient nécessaire d'ajouter des contraintes d'interaction, comme proposé dans [Khan 2004b, Tweed 2002] pour éviter que plusieurs filtres convergent vers la même cible. Il est aussi possible d'incorporer une méthode d'association de cibles de type JPAD (e.g. [Jaward 2006]). On note que dans sa version classique, le filtrage particulière est rapidement gourmand en temps de calcul lorsque le nombre de cibles est devenu élevé.

2.2.4.7 Gestion de la 3D pour le suivi de cibles

- **Suivi 2D** : Suivi de fourmis sur un plan à partir d'une vidéo ;

- **Suivi 3D** : Suivi d'oiseaux en vol à partir d'observations 3D.

Les méthodes de suivi visent à estimer l'état d'une cible sous forme d'un vecteur à n -dimensions. Il est donc théoriquement simple de proposer un modèle adapté au suivi 3D d'un objet. En revanche, dans le cas d'approches de type "track-before-detect", il est nécessaire de prendre certaines précautions. En effet, chaque ajout de nouvelle dimension augmente considérablement l'espace à explorer. Cela nécessite alors l'utilisation d'un nombre plus important de particules, ce qui entraîne un surcoût au niveau des calculs. On note aussi qu'avec un grand nombre de dimensions, il peut être difficile de définir la confiance (au bruit) associée à chaque dimension du vecteur d'état et du vecteur d'observation. C'est par exemple le cas des trajectoires d'abeilles. On peut se poser la question d'ajouter ou non une dimension pour laquelle nous ne sommes pas sûr d'avoir des données. Dans nos travaux, nous choisissons d'utiliser toute l'information disponible et d'adapter le système pour qu'il puisse gérer l'information manquante.

2.2.5 Association de cibles pour le suivi par filtrage de Kalman

Nous avons vu dans les sections précédentes différentes méthodes de pistage permettant d'estimer la position d'une cible à partir d'une séquence de mesures bruitées. Le problème se complexifie lorsque l'on souhaite suivre plusieurs cibles simultanément, notamment dû aux croisements et aux éventuelles occultations prolongées. La difficulté est d'autant plus élevée si les cibles sont susceptibles d'entrer/quitter la scène, ou d'apparaître/disparaître au milieu de la scène. La littérature fait référence à différentes méthodes dites "d'association de cibles" dont l'objectif est de traiter ces problèmes. Dans cette section, nous nous intéressons à ces méthodes, et en particulier à celles compatibles avec les approches de pistage par filtrage de Kalman. La figure 2.5 illustre différentes approches d'association de cibles que nous détaillerons par la suite.

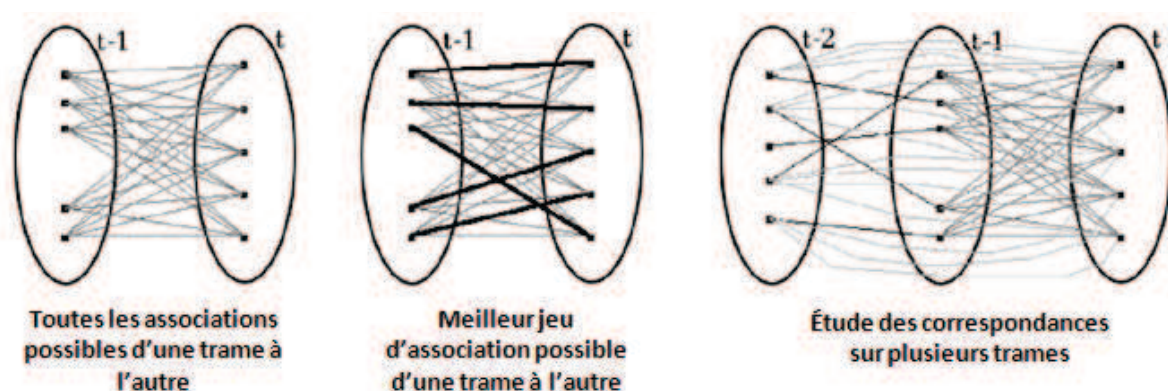


Figure 2.5 – Illustration des différentes méthodes d'association de cibles, naïves (à gauche) et plus évoluées (à droite). Figure adaptée de [Yilmaz 2006].

2.2.5.1 Global Nearest Neighbors (GNN)

GNN (pour Global Nearest Neighbors) [Blackman 1999] est la méthode la plus communément utilisée en termes d'association de "traces/observations". Cette méthode a l'avantage d'être relativement peu complexe et d'être plutôt simple à implémenter. Elle permet de définir, à chaque itération du suivi, la meilleure configuration d'association "traces/observations" ainsi que de gérer la création de nouvelles traces et la destruction de ces dernières. De manière formelle, la matrice d'affectation $[A[c_{i,j}]]$ comprend tous les coûts possibles correspondant aux trois situations citées précédemment. A inclut la possibilité pour chaque observation d'être :

- associée à une trace existante ;
- de n'être associée à aucune trace (en cas de fausse alarme) ;
- de donner naissance à une nouvelle trace.

$c_{i,j}$ est le coût pour l'observation i d'être associée à la possibilité j . De manière globale, la meilleure solution est celle qui minimise la somme des coûts des associations. On pose alors le problème linéaire suivant, pouvant être résolu par l'algorithme Hungarian [Kuhn 1955].

$$\begin{aligned} \sum c_{ij}^2 \cdot x_{ij} \quad & x_{ij} \in \{0, 1\} \\ \sum_i x_{ij} = 1 \quad & \sum_j x_{ij} = 1 \end{aligned} \quad (2.9)$$

En reprenant les équations du filtre de Kalman (eq. 2.5), la distance de Mahalanobis d^2 entre une observation et une trace est définie par :

$$\begin{aligned} d^2 &= (Z - H\hat{X}^-)' S^{-1} (Z - H\hat{X}^-) \\ \text{avec } S &= H P^- H' + R \end{aligned} \quad (2.10)$$

où $(Z - H\hat{X}^-)$ est l'innovation, avec \hat{X}^- l'état prédit *a priori*, Z la mesure et H le modèle d'observation liant l'espace d'observation à l'espace d'état. S est la matrice de covariance de l'innovation avec R la matrice de bruit de mesure, P^- la matrice de covariance d'erreur prédite *a priori*. On cherche ensuite à calculer le coût d'association $c_{i,j}^A$ entre une observation et une trace existante en prenant en compte l'incertitude de l'état estimé (via C_K) et l'incertitude de mesure (via C_S), ce qui donne la relation :

$$\begin{aligned} C_{i,j}^A &= C_K + C_S \\ C_K &= \ln(\sqrt{|S|}) - \frac{N \ln(2\pi) + d^2}{2} \\ C_S &= \ln\left(\frac{P_D}{1-P_D}\right) \end{aligned} \quad (2.11)$$

où P_D est la probabilité que l'observation corresponde à une cible réelle, et N représente la dimension du vecteur d'état. Les coûts $C_{i,j}^{FA}$ et $C_{i,j}^{NT}$, correspondant respectivement à une situation de fausse alarme ou de nouvelle trace, sont donnés par :

$$\begin{aligned} C_{i,j}^{FA} &= -\ln(dC) \\ C_{i,j}^{NT} &= -\ln(dN) \end{aligned} \quad (2.12)$$

où dC et dN correspondent respectivement aux fonctions de densité relatives à l'espace d'observation de la scène. Par exemple, dN peut être modélisé par une carte où chaque position (ou point) est pondérée par sa distance au point d'apparition potentielle de la cible le plus proche.

Finalement, les observations associées à des traces sont traitées par le filtre de Kalman via le cycle de correction/prédiction. Les observations non associées deviennent candidates

pour l'initialisation de nouvelles traces. Une trace est détruite si elle n'est associée à aucune observation pendant un laps de temps limité. Plus de détails sur GNN sont données dans [Blackman 1999].

2.2.5.2 Multi Hypothèse Tracker (MHT)

Contrairement à GNN qui détermine localement la meilleure hypothèse d'association "traces/observations", le MHT étudie la meilleure configuration d'association sur un ensemble de trames successives, et détermine ainsi la meilleure solution de manière différée. Les différentes hypothèses sont stockées dans un arbre de décision. L'arbre grandit à chaque nouvelle itération du suivi, ce qui n'est pas sans poser de problème. Différentes astuces indispensables à l'implémentation sont mises en œuvre afin d'éviter l'exposition combinatoire. La première astuce est l'élagage (ou pruning), permettant de limiter l'arbre à une taille raisonnable (e.g. profondeur correspondant à 3 itérations). La seconde astuce s'appuie sur l'algorithme de Murthy [Miller 1997] pour conserver seulement un nombre limité des meilleures hypothèses à chaque itération. Finalement, un mécanisme de fusion vérifie et supprime les traces similaires susceptibles de suivre une même cible.

En ce qui concerne le score des hypothèses, ce dernier est calculé par la somme de tous les coûts d'association des traces concernées par l'hypothèse depuis le début de chaque trace. Le MHT partage la même base que GNN pour le calcul des coûts d'association "traces/observations" (cf. Section 2.2.5.1). Plus de détails sur MHT sont données dans [Reid 1979].

L'état de l'art déroulé dans ce Chapitre 2 fait référence à un large panel de méthodes capables de répondre à des problèmes divers et variés en termes de filtrage et de pistage multi-cibles. Cependant, on note un certain manque quant à d'éventuels travaux susceptibles de solutionner pleinement la problématique que nous avons formulée en début de chapitre. Notamment, en ce qui concerne le suivi de cibles en 3D à partir d'observations mixtes (2D/3D), ou bien la prise en compte d'une dynamique de cible contextualisée par rapport à la scène. La section suivante aborde ces problèmes spécifiques et propose plusieurs solutions.

2.3 Solutions théoriques proposées

Suite au constat effectué dans la section précédente d'un manque de méthodes clés en main applicables dans le cadre de notre application, nous proposons dans cette section des solutions plus adaptées à notre problème. Celles-ci viennent étendre le champs d'applications des approches classiques de suivi multi-cibles telles que, le filtrage de Kalman, GNN et MHT. Nous avons classé nos propositions en trois contributions distinctes.

Notre première contribution, à laquelle nous ferons référence sous l'appellation "Contrib. 1" (cf. Section 2.3.1), est une méthode permettant lors du suivi d'une cible, de gérer temporairement les observations incomplètes (e.g. profondeur manquante), en projetant l'espace d'observation et de suivi dans un espace 2D.

Nos deux contributions suivantes (cf. sections 2.3.2 et 2.3.3), auxquelles nous ferons référence respectivement sous l'appellation "Contrib. 2a" et "Contrib. 2b" répondent au problème suivant : *"comment prendre en compte l'influence de la scène sur la dynamique des cibles, dans le but d'améliorer le suivi dans le cas multi-cibles ?"* Ces deux contributions apportent chacune indépendamment l'une de l'autre une réponse viable au problème, mais suivent deux

approches fondamentalement différentes. La première approche proposée (Contrib. 2a) s'appuie sur l'injection d'un *a priori* fort sur le modèle de la scène permettant ainsi de contraindre l'estimation de la trajectoire au niveau du filtre de Kalman. La deuxième approche propose (Contrib. 2b) d'intégrer les mêmes contraintes environnementales mais cette fois-ci sans injection directe d'un *a priori* en s'appuyant principalement sur les données.

2.3.1 Filtrage de Kalman 3D et mesures 2D/3D (Contrib. 1)

En ce qui concerne les applications de vidéosurveillance, les observations coordonnées sont généralement exprimées dans le repère image. En revanche, la dynamique des cibles est généralement exprimée dans le repère du monde réel sur la base de grandeurs physiques (e.g. vitesse en m/sec). Ainsi, dans de nombreuses applications s'appuyant sur le filtre de Kalman, il est préférable d'estimer l'état des cibles (e.g. position, vitesse) et des autres modèles (e.g. observation, évolution, bruit) dans le monde réel plutôt que simplement dans l'espace d'observation donné par le capteur (i.e. repère image). Par conséquent, avant de fournir les mesures en entrée du filtre, il est nécessaire de projeter celles-ci dans le repère du monde réel.

La difficulté est que, dans les applications de suivi de cibles en 3D, les limitations technologiques des capteurs ne permettent pas de fournir en permanence des mesures définies en 3D en entrée du filtre. Prenons l'exemple d'un capteur de type "caméra stéréovision", ce dernier, comme un bon nombre de capteurs 3D (e.g. Microsoft Kinect), ne permet pas à coup sûr de fournir la profondeur associée à une observation. Cela pose évidemment problème. Dans ce contexte, nous présentons dans la section 2.3.1.1 le cas "sans problème" où l'information est toujours disponible, puis dans la section 2.3.1.2 nous abordons une solution pour les cas plus problématiques où l'information de profondeur est parfois manquante.

2.3.1.1 Cas optimal : observations projetées dans le monde réel

Dans le cas optimal, la carte de profondeur fournie par la caméra stéréovision renseigne, pour chaque pixel de l'image d'intensité, la distance par rapport au capteur. Ainsi, à chaque observation de cible, on peut associer une position (u, v, d) exprimée dans le repère 2D de l'image d'intensité, où u et v sont les index de colonne et de ligne, et d est la profondeur calculée en ce point. L'équation de projection (eq. 2.13) permet alors d'exprimer les observations dans le repère du monde réel plutôt que dans le repère image. La figure 2.6 illustre le passage d'un repère à l'autre.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = d \begin{bmatrix} \frac{u - cu}{f_u} \\ \frac{v - cv}{f_v} \\ 1 \end{bmatrix} \quad (2.13)$$

avec (cu, cv) , les paramètres de calibration de la caméra stéréovision, se référant aux coordonnées en pixel du point principal, et (f_u, f_v) les distances focales en pixels le long des axes respectifs x et y .

En théorie, dans le cas optimal, le suivi en 3D de plusieurs cibles à l'aide du filtre de Kalman est plutôt efficace. En effet, comparé à la 2D, la 3D permet de réduire considérablement les ambiguïtés d'association lors des croisements de cibles. Par exemple, grâce à l'appréhension de la scène en 3D, il est aussi plus aisé de gérer les apparitions/disparitions de cibles autour de zones définies comme candidates (e.g. obstacle modélisé en 3D au milieu de la scène).

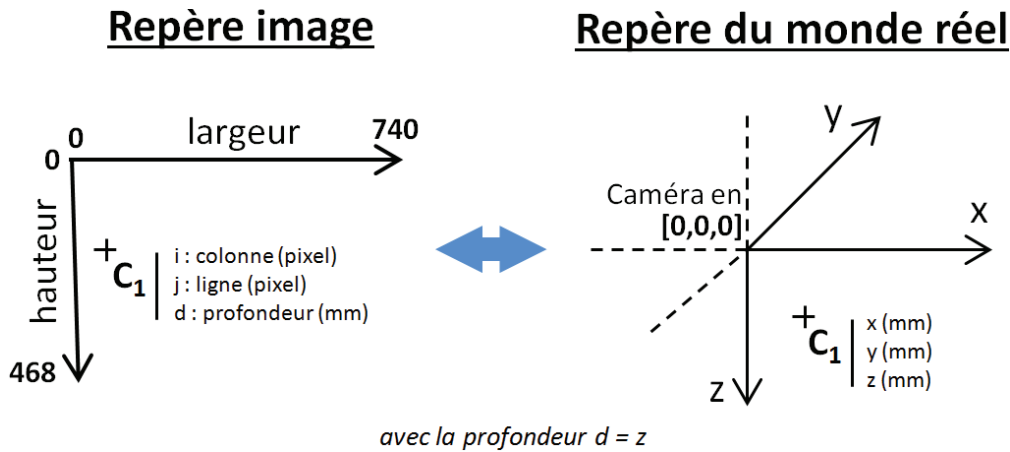


Figure 2.6 – Observation exprimée dans le repère image et dans le repère du monde réel.

2.3.1.2 Gestion de l'absence d'information de profondeur

Contrairement au cas optimal, la réalité veut que l'information de profondeur ne soit pas toujours disponible. Les raisons peuvent être multiples : limite d'exploitation du capteur, illusion d'optique, information jugée non fiable. Dans ce cas, il est alors impossible de projeter l'observation dans le repère du monde réel, et par conséquent cela rend sa prise en compte par le filtre de Kalman impossible.

De plus, il est difficilement envisageable de calculer la profondeur à partir d'autres valeurs observées (e.g. taille de l'ellipse englobante, cf. Section 1.4.3). Afin d'apporter une réponse à ce problème, nous proposons l'approche suivante rendant possible l'intégration des observations dépourvues d'information de profondeur dans la boucle de "prédiction/correction" du filtre de Kalman.

Lorsque la profondeur est inconnue, la valeur de z prédite par le filtre de Kalman est temporairement utilisée comme valeur estimée *a priori* de d . Cela permet de projeter l'observation en question dans le monde réel via l'équation (eq. 2.13). Néanmoins, afin d'éviter une propagation importante de l'erreur, le nombre d'estimations de z (ou d'itération) utilisées pour la projection est volontairement limité. Lorsque l'information de profondeur devient à nouveau disponible au niveau de la séquence d'observations, la série des valeurs de profondeur manquantes est alors recalculée par interpolation (i.e. méthode des splines). Ainsi, toutes les observations dont la profondeur d avait été estimée "en ligne" grâce à l'information de profondeur *a priori* sont maintenant re-projetées sur la base de la série des valeurs de profondeur z interpolées. La trace concernée par ces observations est alors re-estimée (depuis l'étape où l'on a noté l'absence de profondeur) via le processus de "prédiction/correction" classique du filtre de Kalman défini en 3D. Cela permet faire coller l'estimation des états successifs au mieux aux données observées et aussi d'améliorer l'estimation de la prédiction, ce qui va permettre d'augmenter les chances d'association avec une prochaine observation. L'algorithme 1 formalise le mécanisme décrit. Les figures 2.7 et 2.8 illustrent respectivement le mécanisme local de projection en cas d'absence d'information de profondeur, et le mécanisme de reprojektion dans sa globalité appliqué suite à l'absence d'information de profondeur pendant un laps de temps donné.

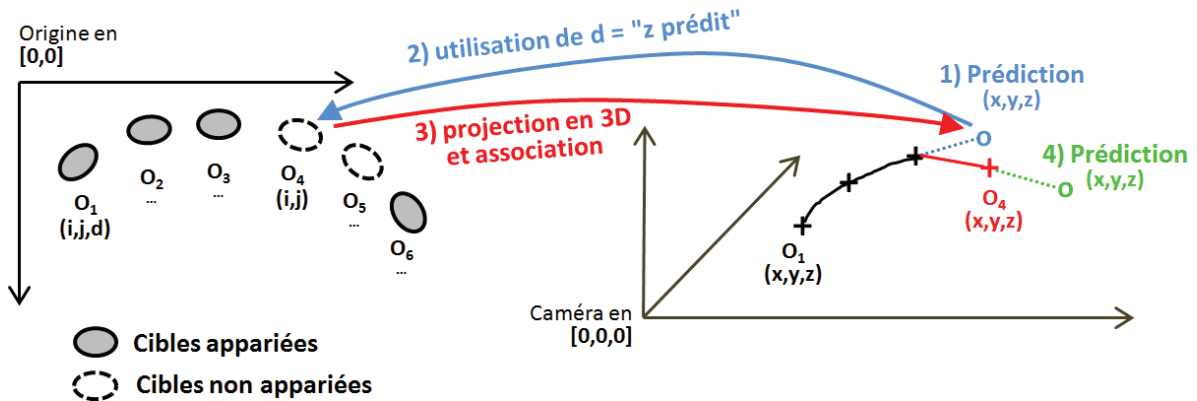
Repère image**Repère du monde réel**

Figure 2.7 – Mécanisme de projection local en cas d'absence d'information de profondeur. Les étapes dans l'ordre d'exécution sont : 1) prédiction de la position (x,y,z) de la cible à l'itération suivante ; 2) utilisation du z prédit pour renseigner la profondeur de la cible ; 3) projection de la cible dans l'espace 3D et association au filtre de Kalman pour l'étape de correction ; 4) bouclage en retournant à la situation 1).

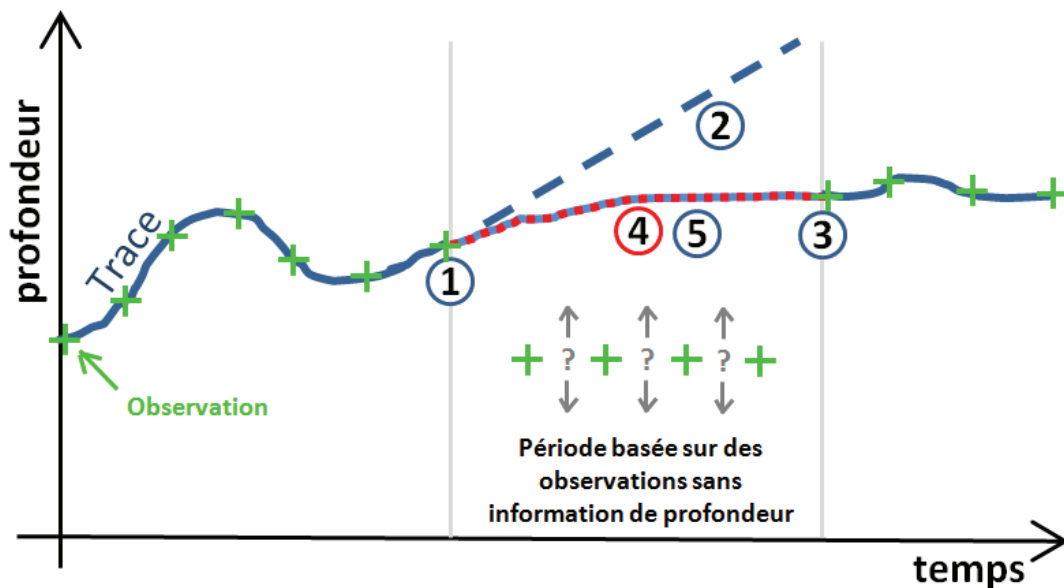


Figure 2.8 – Mécanisme de reprojection suite à l'absence d'information de profondeur pendant un laps de temps donné. Les étapes dans l'ordre sont : 1) début de l'absence de profondeur ; 2) projection de l'observation sur la base du z prédit par le filtre de Kalman ; 3) information de profondeur disponible à nouveau ; 4) interpolation des valeurs de profondeur z pour chaque pas de temps estimé sans profondeur ; 5) re-estimation par le filtre de Kalman de la trace sur la base des nouvelles valeurs de z calculées.

```

while Observation available for a target do
  1) Prediction
  2) Observation
  if  $z$  missing then
    % Enter in a temporary estimation loop
    while  $Z$  missing do
      a) Projection (using the last predicted  $z$ )
      b) Update
      c) Prediction
      d) Observation
    end
    Interpolate  $z$  from the last and new observed  $z$ 
    Reproject observations with new interpolated  $z$ 
    Reapply the KF with new reprojected observations
  end
  3) Projection
  4) Update
end

```

Algorithme 1 : Mécanisme de reprojexion pour la gestion des observations sans profondeur

Une trace peut être initialisée avec une profondeur inconnue. Dans ce cas, une profondeur arbitraire (e.g. 36 cm) est utilisé. Ainsi, dès qu'une observation avec information de profondeur est associée à la trace, toutes les observations précédemment associées à la trace sont re-projetées en fonction la nouvelle profondeur, et la trace est re-estimée entièrement à partir de ces nouvelles observations.

2.3.2 Suivi contraint par un modèle de scène (Contrib. 2a)

Les méthodes d'association de cibles telles que GNN ou MHT s'appuient sur des comparaisons de coûts. C'est à dire que, si le coût d'association "observation/trace" $C_{i,j}^A$ (eq. 2.11) est plus élevé que, soit le coût de "fausse alarme" $C_{i,j}^{FA}$ (eq. 2.12) ou bien le coût de "nouvelle trace" $C_{i,j}^{NT}$ (eq. 2.12), la trace peut n'être associée à aucune observation et peut être à terme éventuellement perdue. Nous appelons ce seuil d'association ACL (pour Association Cost Limitation) défini par le plus petit des coûts parmi $C_{i,j}^{FA}$ et $C_{i,j}^{NT}$.

Lorsque que l'on cherche à suivre des cibles ayant une dynamique erratique, il est préférable de définir un ACL relativement élevé de manière à tolérer une grande distance entre l'état prédit et l'observation, sans quoi la trace serait rapidement perdue. Dans certaines applications de suivi de cibles et notamment celles en 3D, il n'est pas rare d'observer une activité plutôt concentrée sur certaines zones d'intérêt. Cela pose problème car une grande tolérance au niveau de l'ACL peut alors entraîner un grand nombre de mauvaises associations. Ce phénomène est par ailleurs amplifié par une présence potentiellement plus élevée de fausses alarmes dans la zone (due à l'activité importante). De plus il n'est pas rare que la zone d'intérêt en question soit plane, et que les cibles aient tendance à évoluer sur un même plan, rendant alors l'intérêt de la 3D limité. Et finalement, si par malchance les zones d'intérêts correspondent à des zones d'apparition et de disparition de traces, les mauvaises associations peuvent faire perdurer des traces sensées se terminer, et ainsi monopoliser des observations qui devrait plutôt être

associées à de nouvelles traces. Cela rend tout simplement les traces non exploitables par la suite.

Nous partons alors du constat que les coûts d'association "observations/traces" ne doivent pas être uniformes dans la scène, et donc doivent dépendre de la position ou du contexte. La contribution présentée dans cette section s'appuie alors sur l'hypothèse suivante : les zones d'intérêt (souvent très peuplées) font état de ralentissements contraignant le déplacement des cibles, et par conséquent rendent localement les mouvements moins chaotiques. Nous proposons donc de prendre en compte la distance entre les cibles et les zones d'intérêts potentielles dans le calcul des coûts d'association. Cela devrait améliorer la qualité de la prédiction par le filtre de Kalman et ainsi limiter les mauvaises associations.

Le calcul du coût d'association "observation/trace" s'appuie sur une distance de Mahalanobis d^2 (eq. 2.10) qui prend en compte l'incertitude associée à l'estimation de la trace et à l'observation. Ainsi, nous proposons de jouer sur les incertitudes associées à l'estimation de la trace pour adapter la distance d^2 de manière à normaliser (en fonction de la distance à la zone d'intérêt) les valeurs de coût par rapport au seuil ACL fixé en dur. La figure 2.9 illustre le principe.

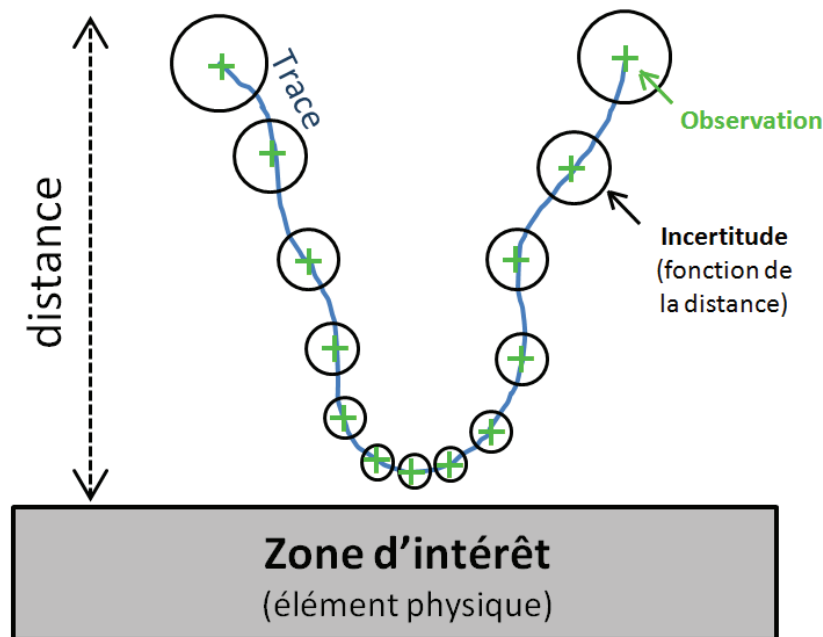


Figure 2.9 – Estimation du bruit l'état contraint par une zone d'intérêt de la scène.

Dans [Chiron 2013b], nous avons proposé une solution qui s'applique bien à l'application de suivi des abeilles. Dans ces travaux, la normalisation du coût d'association d^2 s'appuie sur des coefficients définis en fonction de la distance des cibles (i.e. abeilles) à la zone d'intérêt (e.g. planche d'envol). Ces coefficients sont calculés à partir d'une base d'observations conséquente. Dans les deux sections suivantes, l'idée est la même, mais la pseudo normalisation se fait de manière plus élégante et surtout s'intègre bien mieux au sein des équations du filtre de Kalman.

2.3.2.1 Incertitude P_0 à la naissance des traces

Cette section détaille, comment de manière intuitive, en s'appuyant sur les équations du filtre de Kalman originales, il est possible d'adapter l'incertitude d'estimation liée à de nouvelles traces en fonction de la distance à la zone d'intérêt.

Traces instanciées à proximité de la zone d'intérêt À proximité de la zone d'intérêt, en dépit d'une activité importante, nous faisons l'hypothèse que la dynamique des cibles est relativement calme. On peut donc se permettre de fixer une incertitude initiale P_0 plutôt faible. Cela aura pour effet d'augmenter la valeur de d^2 et ainsi contraindre fortement l'écart par rapport au seuil ACL fixé. En conséquence, les observations candidates pour association à la trace seront alors choisies dans un rayon plus restreint.

Traces instanciées loin de la zone d'intérêt En marge de la zone d'intérêt (e.g. à la limite du champ de vision de la caméra), la dynamique potentiellement rapide et chaotique des cibles nécessite, dès l'origine de la trace, de fixer une incertitude initiale P_0 élevée. Cela offre une plus grande tolérance par rapport au seuil ACL fixé.

Dans les deux cas, que P_0 soit faible ou élevé, l'incertitude P va converger (à la baisse ou à la hausse) vers un état stable suivant le principe de convergence du filtre de Kalman. Finalement, suivant cette approche, les traces au départ d'une zone d'intérêt seront moins vulnérables à de mauvaises associations, et les traces entrant dans la scène seront plus aptes à être associées à des observations lointaines. Ce moyen d'action reste cependant limitée. En effet, la convergence du filtre de Kalman est fonction du nombre d'itérations, alors que dans certains cas, il serait plus pertinent de lier l'incertitude P à d'autres critères, par exemple spatiales (e.g. vitesse de la cible, distance par rapport à une zone d'intérêt).

2.3.2.2 Incertitude P influencée par la distance à la zone d'intérêt

Nous souhaitons lors du suivi, établir un lien permanent entre l'incertitude P et la distance de la cible à la zone d'intérêt. Dans cette optique, nous proposons d'adapter les équations du filtre de Kalman de manière à rendre l'incertitude P fonction de d_{ZI} , d_{ZI} étant la distance entre la cible et la zone d'intérêt. La manière la plus intuitive est de pondérer la matrice P par la distance d_{ZI} :

$$\begin{aligned} w &= d_{ZI}/d_{max} \\ P_k^- &= (AP_{k-1}^+ A^T + Q)w \end{aligned} \tag{2.14}$$

avec d_{ZI} contraint à être dans l'intervalle $[d_{min}, d_{max}]$, d_{min} et d_{max} étant les distances minimum et maximum d'influence sur l'incertitude. On note que la distance entre la cible et la zone d'intérêt ne peut être calculée que si l'on a la connaissance de la structure (e.g. surface 3D) de la zone d'intérêt.

Finalement, nous verrons dans la section 2.4.3 comment cette contribution (Contrib. 2a) et notamment ses prérequis ont été pris en compte dans le cadre de notre application de suivi d'abeilles.

2.3.3 Suivi contraint par une base de données trajectoires (Contrib. 2b)

A l'issue de la section précédente, nous venons proposer une approche s'appuyant sur la structure de la scène pour la prise en compte de la dynamique des cibles au niveau du modèle d'association "observations/traces". Dans cette section, nous proposons de répondre au même problème mais en suivant une approche fondamentalement différente. L'idée est cette fois-ci de prendre en compte un minimum *d'a priori*, et notamment d'éviter, 1) de poser des hypothèses fortes sur la dynamique des cibles et, 2) d'avoir à définir et à modéliser des zones d'intérêt.

Cette nouvelle approche s'appuie essentiellement sur les données. Elle est particulièrement adaptée à des applications de suivi où l'on a une collecte massive de trajectoires (e.g. colonie d'insectes ou petits animaux). Le but étant de contraindre le suivi des cibles à l'aide d'une base de trajectoires établie en amont. Celle-ci, nécessairement constituée de plusieurs milliers de trajectoires, permet de contenir un grand nombre de dynamiques possibles de cibles dans un même contexte de scène.

2.3.3.1 Construction de la base données des trajectoires

Dans un premier temps, nous devons établir une base de données dite grossière rassemblant des trajectoires représentatives de la dynamique de la colonie. Étant donnée que nous n'utilisons pas de connaissance *a priori* sur la scène, les trajectoires pour cette base de données sont collectées suivant une configuration de suivi relativement permissive. Naturellement, cela engendre des trajectoires incomplètes ou partiellement fausses à cause aux erreurs classiques de suivi telles que : des cibles non suivies, des mélanges de traces ou des traces terminées prématurément. Mais même avec un taux d'erreur élevé, les chances pour que plusieurs erreurs se produisent au même endroit dans l'espace 3D sont relativement faibles. On peut dire que la base de données de trajectoires représente globalement un grand nombre de mouvements possibles en tout point de l'espace. Notamment, nous considérons que l'importante quantité de données compense les potentielles erreurs que contient la base. Finalement, l'intérêt de cette base est qu'elle représente la dynamique des possibles, et contient par conséquent intrinsèquement la structure contrainte de la scène.

2.3.3.2 Prédiction du KF basé sur la base de données de trajectoires

Notre solution consiste à remplacer la prédiction du filtre de Kalman afin d'estimer l'état directement à partir des dynamiques enregistrées dans la base de données de trajectoires. De manière formelle, posons Q comme étant une sous-trace dite "requête" pour laquelle nous souhaitons prédire l'état à venir. $DB_{trajectories}$ est la base de données des trajectoires dont la construction a été décrite dans la section précédente. $DB_{subTracks}$ est l'ensemble des sous-traces de la même taille que Q pouvant être extraites de $DB_{trajectories}$. En utilisant une fonction Euclidienne pondérée pour le calcul de la distance (accordant plus de poids aux points les plus récents des traces), un ensemble C de n sous-traces les plus proches de Q est calculé à partir de $DB_{subTracks}$. Ensuite, en considérant l'ensemble des états à venir de chacune des sous-traces de C , nous estimons l'état suivant de Q . Le processus complet est détaillé au travers des points suivants.

1. Définition d'une sous-trace requête Q d'une durée l correspondant à une trace TC en cours de suivi dont l'état a déjà été estimé à partir du début de la trace $t = 1$ jusqu'à

$k - 1$:

$$Q \leftarrow TC_{k-l:k} \quad \text{with} \quad l \leftarrow \min(l_Q, TC_{length}) \quad (2.15)$$

avec l_Q la durée maximum définie pour une requête (e.g. 15 pas de temps).

2. Extraction de la base de données de sous-traces $DB_{subTracks}$ à partir de la base de données de trajectoires $DB_{trajectories}$:

```

Data :  $DB_{subTracks} = \emptyset$ 
for  $T \in DB_{trajectories}$  having  $T_{size} > l$  do
  | for  $i \leftarrow 1$  to  $T_{size} - l$  do
  | | add  $T_{i \rightarrow i+l}$  in  $DB_{subTracks}$ 
  | end
end

```

3. Trouver à partir de $DB_{subTracks}$, l'ensemble C des n sous-traces les plus proches de Q (la sous-trace requête), ou plus spécifiquement les index correspondants :

$$C\text{-indexes} \leftarrow \text{sort}(\text{dist}(Q, X), \forall X \in DB_{subTracks})_{1:n} \quad (2.16)$$

using $\text{dist}(Q, X) = \sum_{i=1}^l \omega_i (X_{(i,d)} - Q_{(i,d)})^2$

avec ω proportionnel à i par exemple, et i étant l'indice du point (coordonnées exprimées en d dimensions) de la trace.

4. Remplacement de la phase de prédiction du filtre de Kalman. L'état moyen et l'incertitude associée (matrice de covariance) sont données par :

$$\hat{x}_k^- = [\mu_x, \mu_y, \mu_z, \dots]^T \quad (2.17)$$

$$\hat{P}_{k(m,n)}^- = \text{cov}(C\text{-coords}_{(m)}, C\text{-coords}_{(n)})$$

avec $C\text{-coords}_{(d)}$ l'ensemble des coordonnées pour chaque point correspondant à l'état suivant de chaque sous-trace de C , selon la dimension d , à savoir (x, y, z, \dots) . L'état moyen est obtenu par $\mu_d = \mathbb{E}[C\text{-coords}_{(d)}]$. Il faut noter que nous appliquons au préalable sur chaque point de $C\text{-coords}$ une translation pour recentrer le mouvement potentiel dans le contexte de la requête Q .

$$C\text{-coords}_{(i,d)} = T_{(i,d)_{l+1}} - (T_{(i,d)_l} - Q_{l,d}) \quad (2.18)$$

avec $T_{(i,d)} = DB_{subTracks}(C\text{-indexes}_{(i)})$

avec i l'indice de la trace et l l'indice du point dans la trace, et d la dimension des coordonnées (i.e. 2D ou 3D). La figure 2.10 illustre l'algorithme présenté ci-dessus.

Les chances de succès de cette approche sont étroitement liées au nombre de trajectoires collectées. Les applications qui offrent une grande liberté de mouvement au niveau des cibles nécessiteront naturellement une plus grande base de trajectoires que les applications où les

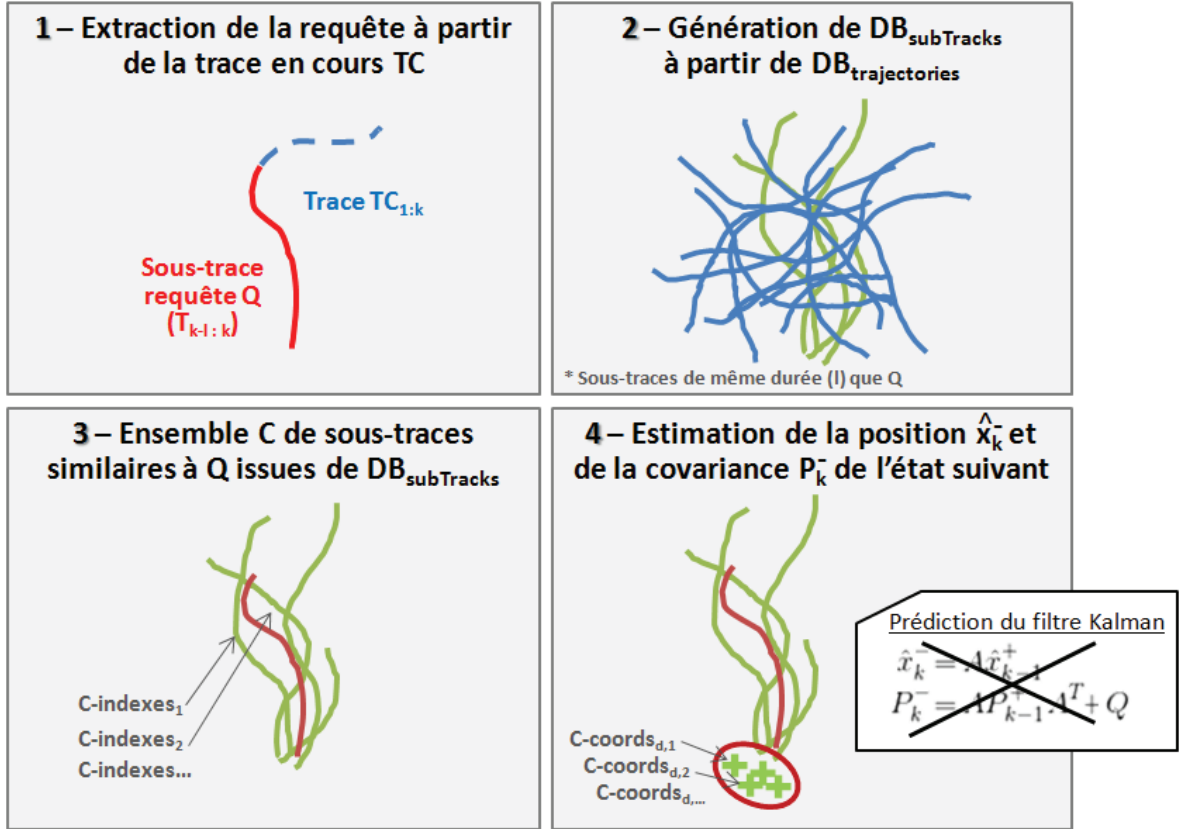


Figure 2.10 – Illustration des différentes étapes du processus de prédiction effectué sur la base de données de trajectoires.

mouvements des cibles sont fortement contraints, voire canalisés. À partir des expérimentations effectuées dans [Chiron 2014a], nous estimons comme ordre de grandeur à plusieurs milliers le nombre de trajectoires nécessaires dans le cadre d'une application présentant peu de contraintes. À titre indicatif, la figure 2.11 montre que dans une situation hypothétique de suivi en 2D, une base de données d'au moins 25k trajectoires est nécessaire pour atteindre une valeur d'erreur sur la prédiction de l'état comparable aux méthodes classiques de prédiction. Et naturellement, avec l'ajout d'une dimension supplémentaire, il semblerait difficile d'atteindre des performances équivalentes. Il faudrait probablement une base de données bien plus conséquente. Néanmoins, nous verrons lors de l'évaluation de la méthode que cette augmentation du taux d'erreur n'est pas rédhibitoire. Concernant la prédiction de l'incertitude, une base de données importante aura pour effet de réduire l'incertitude, ce qui apparaît fondamentalement logique.

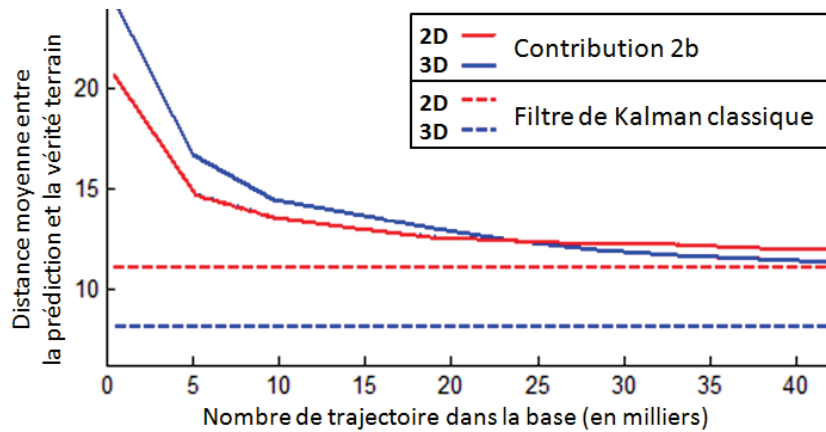


Figure 2.11 – Erreur sur l'état prédit relativement à la taille de la base de données de trajectoires. Les tests ont été réalisés à partir de données issues d'une application mettant en scène des cibles relativement peu contraintes (cf. [Chiron 2014a]). Les trajectoires dites "2D" sont les mêmes que celles définies en 3D, à la différence que l'on ne s'occupe pas de l'information de profondeur.

2.4 Préparation de la mise en œuvre pour le suivi d'abeilles

Dans cette section, nous proposons de mettre à l'épreuve nos différentes contributions (i.e. Contrib 1, 2a et 2b) présentées dans ce chapitre. Pour cela, nous nous focalisons sur l'application phare de cette thèse : le suivi 3D des abeilles en vol devant leur ruche. Dans le chapitre précédent, nous avons présenté notre méthode de segmentation HIDS accompagnée d'un exemple pour la détection d'abeilles en 3D. À présent, l'idée est d'utiliser ces données d'observations pour suivre les abeilles en 3D et reconstituer leur trajectoires individuelles.

Bien que notre méthode de segmentation HIDS soit robuste, la nature de la scène observée (i.e. activité importante, dynamique des cibles) fait que les données d'observations soit complexes, et leur traitement s'avère être un véritable challenge. La section qui suit détaille les difficultés de l'application et souligne notamment l'intérêt de la mise en œuvre de nos contributions.

2.4.1 Présentation de l'application

La figure 2.12 ci-après illustre de manière synthétique les problèmes que nous avons identifiés au sujet de notre application de suivi 3D d'abeilles en vol devant la ruche. Ces problèmes sont détaillés dans les 10 points suivants.

1. **Naissance de nouvelles traces** : comment gérer l'apparition de nouvelles cibles dans le champ de la caméra ? Les cibles peuvent apparaître sur les bords de la scène, à la sortie de la ruche, et même simplement réapparaître à n'importe quel endroit de la scène après avoir été préalablement caché par un élément du décor ou une autre cible.
2. **Imprécision sur les observations** : comment gérer les observations imprécises ? L'étape de segmentation renvoie des mesures avec un certain niveau de précision. La précision peut être relativement faible sur certains paramètres (e.g. profondeur).
3. **Incomplétude des données** : comment gérer la continuité d'une trace lorsque que la cible correspondante n'est pas détectée (i.e. observation manquante). Faut-il décider de

détruire la trace après un certain laps de temps sans observation, et si oui, sous quelles conditions ?

4. **Mesures partielles** : Bien que notre méthode de segmentation permette de détecter les cibles, la plupart du temps en 3D, il est possible qu'une cible soit localisée seulement sur le plan image (i.e. 2D). Comment utiliser ces observations partielles à bon escient ?
5. **Fréquence d'observation instable** : le dispositif d'acquisition capture les images à une fréquence comprise entre 25 et 53 Hz, mais de manière instable. Comment cela doit-il être pris en compte dans le modèle de suivi ?
6. **Destruction des traces** : quand décider qu'une trace doit être détruite ? Sachant que les cibles peuvent apparaître/disparaître n'importe où dans la scène.
7. **Problème de continuité des traces** : lié aux problèmes d'apparition et de destruction de traces. Comment décider si deux traces potentiellement joignables appartiennent à une unique cible ou à plusieurs cibles ?
8. **Problème d'association de cibles** : comment décider quelle observation doit être reliée à quelle trace ?
9. **Observations fausses** : comment gérer les observations issues de bruits (fausses alarmes) pour éviter qu'elles ne donnent naissance et propagent de fausses traces ?
10. **Dynamique évolutive** : comment contextualiser le modèle de dynamiques des cibles afin d'améliorer le suivi ?

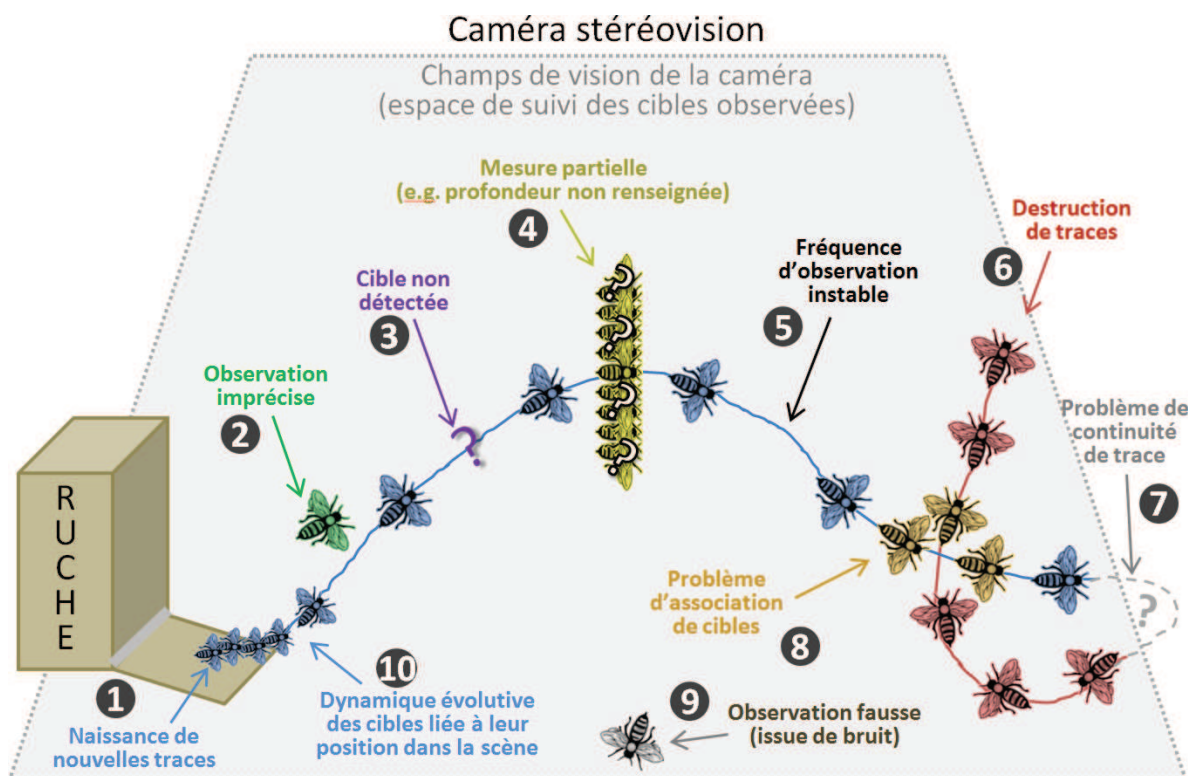


Figure 2.12 – Aperçu des problèmes à résoudre pour la mise en œuvre du suivi d'abeilles en 3D devant la ruche à partir des observations issues de notre segmentation HIDS.

Les auteurs de [Portelli 2011] ont montré l'existence d'une relation entre la vitesse d'une abeille en vol et sa proximité avec les objets qui l'entourent. Nous avons également observé ce phénomène de manière expérimentale devant la ruche en nous appuyant sur les coûts d'association (cf. Figure 2.13). Cela veut dire que la dynamique des abeilles devient plus stable à l'approche de la planche d'envol. A partir de ce constat, nous pensons dans un premier temps qu'il est pertinent de mettre en œuvre la Contrib. 2a dans le cadre de notre application de suivi d'abeilles. Les pré-requis pour la mise en application de cette méthode sont détaillés dans la section 2.4.3.

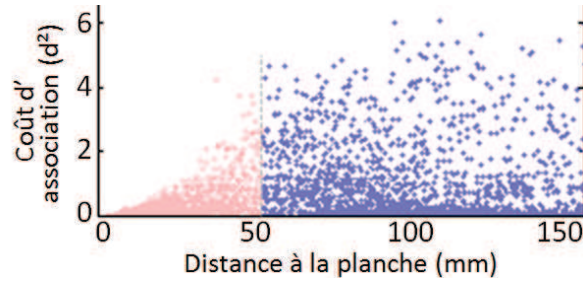


Figure 2.13 – Aperçu, sur un échantillon de traces d'abeilles (obtenues par le couple Filtre de Kalman + GNN), des coûts d'associations par rapport à la distance à la planche d'envol. Données issues de [Chiron 2013b].

Notre dispositif doit permettre de traiter un grand nombre de vidéos. À partir de chaque heure de vidéo, en fonction de l'activité de la colonie, on peut extraire jusqu'à plusieurs milliers de trajectoires. L'approche préconisée en section 2.3.3 (Contrib. 2b) est donc également tout à fait adaptée. Nous pensons alors qu'il est intéressant d'étudier cette piste dans le cadre de notre application. Les pré-requis pour sa mise en application sont détaillés plus tard dans la section 2.4.4. Le tableau 2.1 propose un résumé des problèmes identifiés dans cette section, pour lesquels nous avons proposé des solutions, soit issues de la littérature, soit issues de nos contributions. La mise en œuvre de ces solutions est détaillée par la suite.

Problèmes identifiés	Solutions mises en œuvre
<ul style="list-style-type: none"> - Pistage (Instanciation, destruction et continuité des traces) : 1, 6, 7 - Observations fausses : 9 - Problème d'association : 8 	<ul style="list-style-type: none"> - GNN (2D vs 3D) et - MHT (2D vs 3D) - Confiance sur l'observation définie lors de la segmentation
<ul style="list-style-type: none"> - Cibles non détectées : 3 - Fréquence instable : 5 	- Filtre de Kalman (équations permettant de prédire à un temps donné)
<ul style="list-style-type: none"> - Observations biaisées : 2 	- Filtre de Kalman (modèles de bruit)
<ul style="list-style-type: none"> - Profondeur non disponible : 4 	- Contrib. 1 (cf. Section 2.3.1)
<ul style="list-style-type: none"> - Dynamique évolutive : 10 	- Contrib. 2a (cf. Section 2.3.2) ou - Contrib. 2b (cf. Section 2.3.3)

Tableau 2.1 – Synthèse des problèmes applicatifs identifiés dans cette section 2.4.1, et propositions de solutions, soit issues de la littérature, soit issues de nos contributions (cf. Section 2.3).

2.4.2 Configuration du filtre de Kalman

Soit $Z_{1:n}$ une série d'observation 3D correspondant à une trace au temps 1 à n . Au temps k , une observation est définie en 3D par le vecteur $Z_k = [x, y, z]^T$, et l'état de la cible est estimée par le vecteur $X_k = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$ combinant la position 3D de la cible et sa vitesse.

2.4.3 Modélisation de la planche d'envol

Un pré-requis pour la mise en œuvre de l'approche préconisée en section 2.3.2 est de définir les zones d'intérêt susceptibles d'influencer la dynamique des cibles, et ensuite de modéliser ces zones en 3D. Ces informations sont nécessaires pour pouvoir calculer la distance des cibles par rapport aux zones en question. Dans notre application, on observe une activité intense autour de la planche d'envol, qui est une zone de passage obligatoire pour entrer/sortir de la ruche. Nous définissons donc la planche d'envol comme étant une zone d'intérêt. Il reste alors à la modéliser.

Les images de profondeur calculées par la caméra stéréovision sont bruitées et incomplètes. Prises individuellement en l'état, elles ne permettent pas de modéliser la surface de la planche d'envol de manière satisfaisante. Ainsi nous proposons de calculer le modèle de la planche d'envol à partir d'un ensemble (ou séquence) d'images de profondeur. Ce modèle doit représenter la planche de la manière la plus complète possible. La procédure de calcul doit éviter de prendre en compte les abeilles en vol et celles qui sont posées sur la planche. Paradoxalement, une activité moyenne (en terme de fréquentation) est la condition optimale permettant d'obtenir un ensemble d'images de profondeur efficace en termes de redondance et complémentarité d'information. En effet, le cas d'activité nulle, les zones non texturées (et donc non appariables) ne sont jamais remises en cause par un élément externe (e.g. changements de luminosité, ombres) et resteront indéterminées sur l'ensemble des images de profondeur. À l'opposé, en cas de pleine activité, la planche d'envol peut être couverte d'abeilles et être ainsi complètement occultée. Dans ce contexte, nous proposons ci-après une méthode itérative pour la modélisation de la surface de la planche d'envol. L'objectif étant de représenter le plus fidèlement possible la structure de la surface de la planche d'envol.

Tout d'abord, la région (en 2D) correspondant à la planche d'envol est définie manuellement en superposant un polygone sur une image d'intensité quelconque qui servira par la suite de masque de travail. Ainsi, les traitements suivants sont ensuite effectués sur cette région de manière isolée du reste de l'image. Premièrement, un filtre médian est appliqué sur les images de profondeur une à une de manière à éliminer les erreurs d'appariement engendrées par l'algorithme d'appariement de la caméra stéréovision. Ensuite, un modèle est initialisé à partir de la première image de profondeur de la séquence fournie. Notons que la planche d'envol est représentée par des valeurs de profondeur plus élevées en comparaison des valeurs de profondeur associées aux abeilles qui circulent au dessus celle-ci. Ainsi, de manière itérative, image par image, le modèle est mise à jour en prenant en compte la valeur maximum entre le pixel du modèle et celui de l'image en cours. La figure 2.14 montre étape par étape le raffinement du modèle de la planche d'envol. La difficulté est de trouver le nombre N d'images à utiliser pour la modélisation. N doit être assez élevé pour inclure un nombre d'images suffisant pour assurer une certaine redondance et une bonne complémentarité de l'information, permettant ainsi d'être résistant au bruit. Mais, dû à l'opération de maximisation, une valeur de N trop élevé augmente les chances d'inclure du bruit (e.g. valeurs de profondeur sur-évaluées et non filtrées par le filtre médian de départ). Le nombre d'itérations optimal (ou nombre d'images

N à prendre en compte) peut être estimé en trouvant l'itération qui minimise l'écart-type des valeurs de profondeur du modèle. La figure 2.15 montre un exemple avec $N = 180$, pour une planche d'envol se trouvant en moyenne à 38 cm de l'objectif de la caméra.

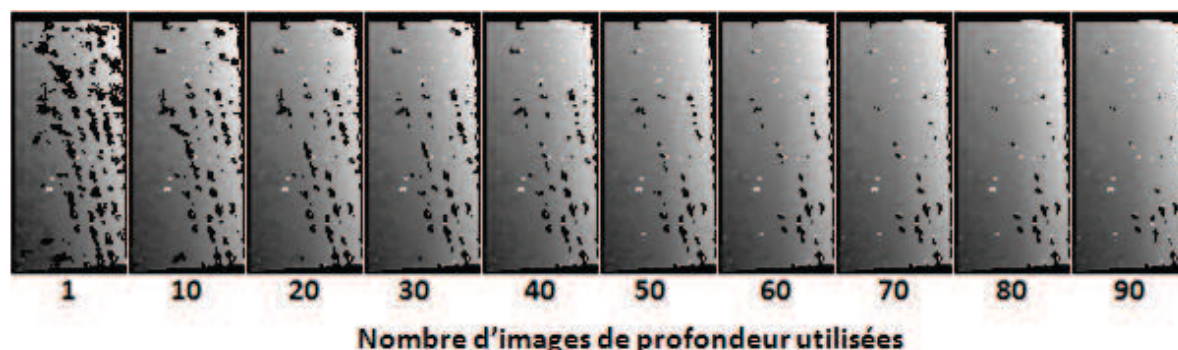


Figure 2.14 – Aperçu de la planche d'envol (vue du dessus, tournée d'un quart de tour) en cours de reconstruction.

Même après ce processus, il se peut que le modèle contienne encore des régions non définies ou incohérentes. Un dernier lissage est donc appliqué sur la base d'une régression LOWESS (Locally Weighted Scatterplot Smoothing) [Cleveland 1981] avec la fonction de Bisquare [Tukey 1962] permettant de gérer les points aberrants. La figure 2.16 montre le maillage 3D reconstruit à partir d'un modèle de surface de planche d'envol légèrement incurvée.

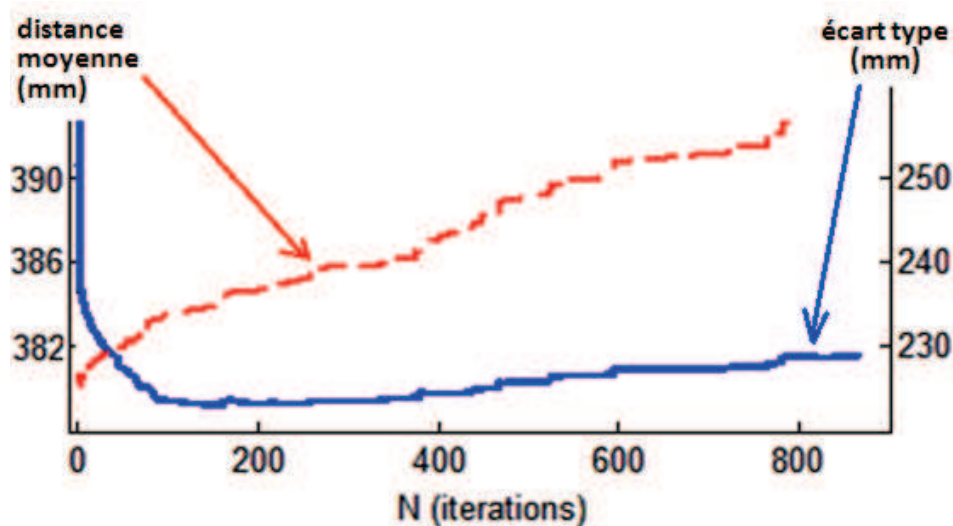


Figure 2.15 – Écart-types (en bleu) et moyennes (en rouge) des valeurs de profondeur de la planche d'envol durant le processus itératif de reconstruction.

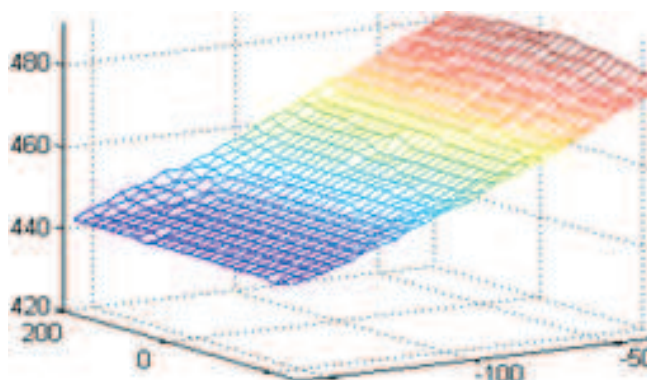


Figure 2.16 – Planche d'envol légèrement incurvée reconstruite sous forme de maillage 3D.

2.4.4 Construction d'une base de données de trajectoires d'abeilles en vol

Un pré-requis pour la mise en œuvre de l'approche de suivi basée principalement sur les données (cf. Contrib. 2b, cf. Section 2.3.3) est de construire une base de données grossière de trajectoires 3D d'abeilles en vol devant la ruche. Pour cela, nous nous appuyons sur un couple "Filtre de Kalman + GNN". Ce couple est paramétré de manière permissive, c'est à dire que le modèle tolère des associations "traces/observations" lointaines. Cela permet de collecter un grand nombre de traces complètes, quitte à surestimer la longueur de certaines traces (e.g. fusions avec d'autres traces). Le but étant d'avoir une base de données représentative de la dynamique des abeilles en vol. La figure 2.17 montre un échantillon des trajectoires issues de la base de données pour l'application de suivi d'abeilles. À titre indicatif, nous utilisons les paramètres suivants : $r=1$ (modèle de bruit de mesure), $q=6$ (modèle de bruit d'évolution d'état) pour filtre Kalman, et pour GNN, les paramètres dC et dN sont doublés par rapport à la solution classique.

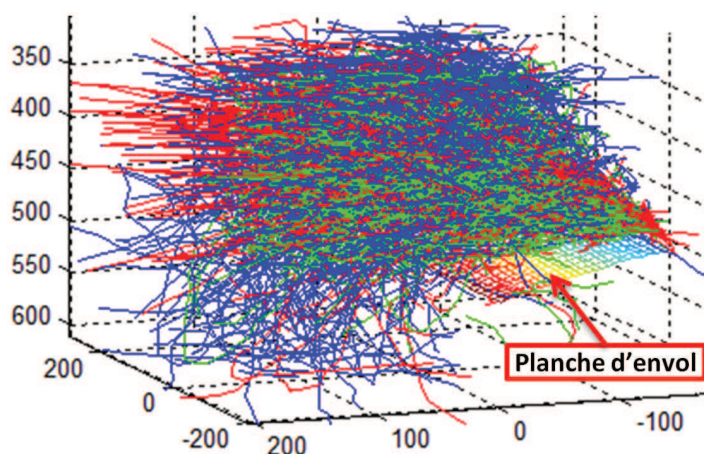


Figure 2.17 – Illustration de 0.1% des trajectoires constituant la base de données. La planche d'envol est affichée sous forme de maillage en guise de référence. À titre indicatif, les couleurs renseignent l'étiquette associée aux traces : rouge pour un départ, vert pour une arrivée, bleu pour une traversée simple sans contact avec la planche d'envol.

2.5 Méthode d'évaluation et résultats

2.5.1 Mise en place d'une vérité terrain

Notre application vise à étudier des milliers de trajectoires d'abeilles en 3D à l'entrée de la ruche. Il serait compliqué et laborieux de créer une vérité terrain de manière exhaustive à l'échelle de l'application. C'est pourquoi nous avons choisi de mettre en place une vérité terrain semi-simulée pour évaluer la performance de nos méthodes de suivi de cibles. Premièrement, nous avons annoté à la main une base de 200 trajectoires en 3D (correspondant à 3640 positions 3D). Ces traces annotées couvrent au total plus de 1000 images. Pour de raisons de ressources humaines limitées et de priorité, nous n'avons pas annoté l'ensemble des cibles présentes sur l'ensemble des images couvertes par les 200 trajectoires, mais seulement les observations d'abeilles correspondant aux traces en question. L'annexe A.2 montre l'interface d'annotation des trajectoires.

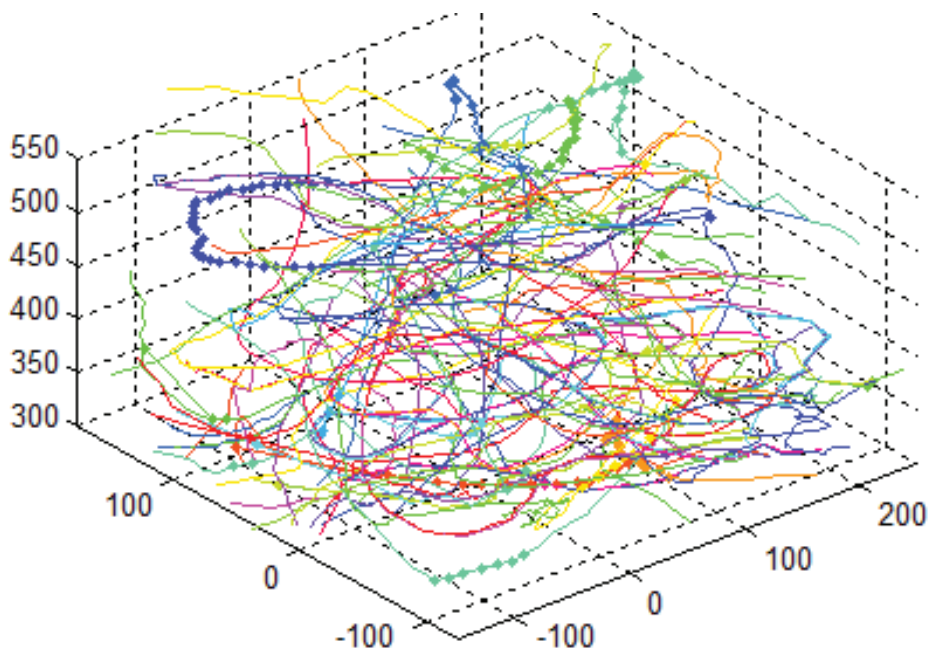


Figure 2.18 – Exemples de trajectoire contenues dans la vérité terrain. Les marqueurs plus épais présents sur les trajectoires indiquent les observations originalement dépourvues d'information de profondeur, pour lesquelles ces informations manquantes ont été interpolées.

À la suite de cette phase d'annotation, nous avons recréé des scénarios virtuels en répliquant et en décalant dans le temps des traces entières tirées aléatoirement à partir de la base des trajectoires annotées. Naturellement, les trajectoires annotées ne sont pas prises en compte dans la base de données de trajectoires définie en section 2.4.4. Pour éviter les confusions, notons que ces deux bases de données (base de trajectoires et vérité terrain) sont complètement différentes.

Cette approche offre la possibilité de créer des scénarios réalistes sur mesure, tout en contrôlant les paramètres basiques tels que le nombre de cibles par exemple. Les fausses alarmes et détections manquées sont également ajoutées aux scénarios en s'appuyant sur

les distributions constatées lors de l'évaluation de la segmentation (cf. Section 1.5.2.2). Une large majorité des fausses alarmes ne sont pas dotées d'information de profondeur, ce qui constitue la difficulté principale du suivi de traces en 3D. En adéquation avec les résultats de l'évaluation de la segmentation, nous fixons à 0.1 la probabilité de connaître la profondeur sur une fausse alarme. Naturellement, les occultations provoquées par les croisements de cibles (i.e. superpositions en vol) sont également pris en compte en retirant dans les scénarios simulés les cibles susceptibles d'être cachées du point de vue de la caméra. Finalement, la probabilité de détection d'une observation est donnée par $P_D = 1 - m_d^{FA}(S \sim \mathcal{N}(\mu_d, \sigma_d))$, avec la fonction d'appartenance m^{FA} définie dans la section 1.4.3, μ_d et σ_d sont les fonctions moyennes et écart type de d , d étant la profondeur d'une cible. En résumé, la vérité terrain générée contient des traces 3D réelles d'abeilles évoluant devant l'entrée de la ruche. L'observation de la scène s'inscrit dans une pyramide de 40×40 cm de base et de 40 cm de hauteur. La figure 2.18 montre la base des 200 trajectoires annotées.

2.5.2 Définition des modèles de scénario

Afin de comparer au mieux les différentes méthodes abordées dans la section 2.3, nous proposons de générer les quatre modèles de scénario ci-dessous. Chacun des modèles met en avant une difficulté particulière rencontrée dans notre application. La configuration basique de ces derniers est définie selon des taux de détections manquées (MD) et de fausses alarmes (FA) fixés respectivement à 11.0% et 28.8%, ce qui correspond à la situation la plus difficile rencontrée lors de l'évaluation de la segmentation (Capture 2, cf. Section 1.5.2.1).

Premier modèle de scénario : "Normal" Modèle de scénario normal (ou témoin) recréant une scène classique d'activité moyenne avec une proportion équivalente de départs, d'arrivées et de vols sans contact (pas d'atterrissage ni de décollage). Ce type de scénario va permettre d'évaluer en condition "normale" les performances de GNN vs MHT, et aussi l'intérêt de l'usage de la 3D par rapport à la 2D.

Deuxième modèle de scénario : "Couple" Modèle de scénario mettant en avant une difficulté majeure de notre application : la gestion des apparitions et disparitions de cibles au niveau de la planche d'envol. Ce scénario va permettre d'évaluer le gain de performance apporté par nos Contrib. 2a et 2b (voir les sections 2.3.2 et 2.3.3) permettant toutes deux d'adapter le suivi en fonction du contexte de la cible par rapport à la scène.

Troisième et quatrième modèle de scénario : "FA+" et "MD+" Ces types de scénario sont similaires au modèle normal, excepté le taux de fausses alarmes (FA) et de détections manquées (MD) qui sont augmentés au niveau des observations. Ces types de scénario vont permettre d'évaluer la robustesse de la méthode retenue dans des conditions extrêmes.

2.5.3 Métrique pour l'évaluation

Il existe différentes métriques pour l'évaluation de la performance d'une méthode de suivi de cibles. Parmi les plus communément utilisées dans la littérature [Yin 2007, Black 2003], on retrouve :

- **Le nombre de traces correctement détectées** ou CDT (pour Correct Detected Track), ayant une superposition spatiale **ET** temporelle suffisante avec une trace de vérité terrain.

$$\frac{\text{Length}(GT_i \cap ST_j)}{\text{Length}(GT_i)} \geq TR_{ov} \quad \text{et} \quad \frac{\sum_{k=1}^N A(GT_{ik}, ST_{jk})}{N} \geq T_{ov} \quad (2.19)$$

- **Le nombre de fausses traces** ou FAT pour (False Alarm Track), ayant une superposition spatiale **OU** temporelle insuffisante avec une trace de VT.

$$\frac{\text{Length}(GT_i \cap ST_j)}{\text{Length}(ST_j)} < TR_{ov} \quad \text{ou} \quad \frac{\sum_{k=1}^N A(GT_{ik}, ST_{jk})}{N} < T_{ov} \quad (2.20)$$

- **Le taux de fragmentation** définie par le rapport entre le nombre de traces CDT (considérées comme correctement détectées) et le nombre de traces de la vérité terrain.

Telle que définie ci-dessus, la métrique CDT est calculée avec une borne supérieure seulement. C'est pourquoi la mesure de fragmentation est bien complémentaire. Dans le contexte de notre application, nous travaillons sur des trajectoires relativement courtes, ce qui rend les traces fragmentées difficilement exploitables. Nous proposons alors de fusionner ces 2 métriques (CDT et fragmentation) en une seule, en bornant totalement CDT temporellement. Nous définissons ainsi notre propre métrique dérivée, appelée TM (pour Track Maintenance). Celle-ci servira de référence pour l'évaluation de la performance de nos méthodes de suivi de cibles.

$$\text{abs}\left(1 - \frac{\text{Length}(GT_i \cap ST_j)}{\text{Length}(GT_i)}\right) \leq TR_{ov} \quad (2.21)$$

À noter que la métrique FAT est finalement peu pertinente dans notre contexte applicatif étant donnée la faible possibilité de générer des traces essentiellement à partir de fausses alarmes. De plus si une mauvaise trace est générée à partir de vraies observations (e.g. fusion ou mélange de 2 traces réelles), cela se traduira automatiquement par une baisse de notre métrique TM. Finalement, nous fixons TR_{ov} à 0.1, ce qui laisse une marge de 10% correspondant à un délai potentiel d'initialisation ou de destruction des traces. T_{ov} est également fixé arbitrairement à 0.1, ce qui laisse une marge acceptable au niveau de l'erreur d'estimation (et filtrage) spatial de la trace.

2.5.4 Résultats sur différents modèles de scénario

Afin d'obtenir des résultats robustes, nous avons lancé chacune des expériences suivantes sur 100 scénarios distincts. Les différents scénarios sont générés aléatoirement à partir des modèles de scénario. Concernant le paramétrage des méthodes de suivi, une distribution similaire à celle observée sur la figure 1.19 est utilisée pour dC . Concernant, dN , étant donné que les cibles n'apparaissent pas seulement sur les bords de la scène mais aussi à l'entrée de la ruche et par les côtés de la planche d'envol, une distribution uniforme est considérée.

2.5.4.1 Dimension d'observation (2D ou 3D) plus importante que le choix de la méthode de suivi (GNN ou MHT)

MODÈLE DE SCÉNARIO UTILISÉ : "Normal")

Dans cette partie, on compare les performances des différentes méthodes ou configurations de suivi s'appuyant sur GNN et MHT. Tel que définie dans la section 2.2.5.2, GNN est un cas particulier de MHT qui garde seulement la meilleure hypothèse à chaque étape du suivi. On note P la profondeur de l'arbre d'hypothèses MHT (niveau d'élagage), et H le nombre de meilleures hypothèses considérées à chaque nœud de l'arbre. On compare alors les trois approches suivantes :

- GNN ($P=1$, $H=1$), configuration naïve, localisée dans le temps,
- MHT faible ($P=2$, $H=3$), configuration avancée, avec une portée temporelle restreinte,
- MHT fort ($P=3$, $H=4$), configuration plus avancée, avec une meilleure portée temporelle.

Aussi, chacune des trois approches est testée sur la base d'observations 3D, ainsi que sur une base d'observations 2D (observations 3D re-projetées sur un plan de profondeur constante).

Premier constat (cf. Figure 2.19) : la complexité de calcul du MHT augmente exponentiellement avec le nombre de cibles, contrairement à GNN qui est moins affecté. Considérant la quantité de données à traiter et l'espoir d'approcher une application temps réel, nous estimons difficile l'utilisation du MHT dans ce cadre.

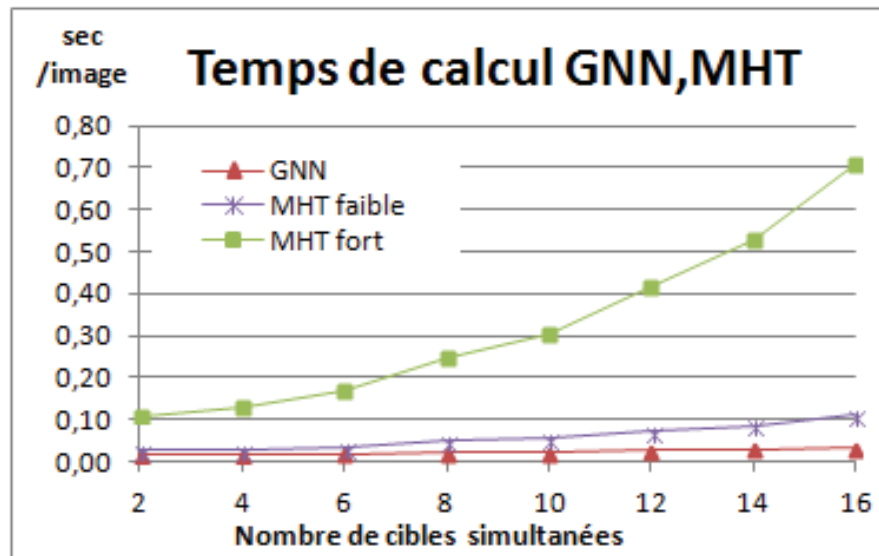


Figure 2.19 – Comparaison des temps de calcul entre les méthodes GNN et MHT. Tests effectués sur une machine équipée d'un CORE i5 à 2.5 GHz.

Deuxième constat (cf. Figure 2.20) : sur nos données, MHT offre une augmentation des performances modérée par rapport à GNN, et ceci spécialement dans les scénarios présentant un nombre de cibles important. Cela peut être expliqué par un nombre de fausses alarmes et de détections manquées relativement bas.

Troisième constat (cf. Figure 2.20) : l'importance de l'information de profondeur pour le suivi de cibles ; la méthode 3D-GNN fournit de meilleurs résultats que la méthode 2D-MHT.

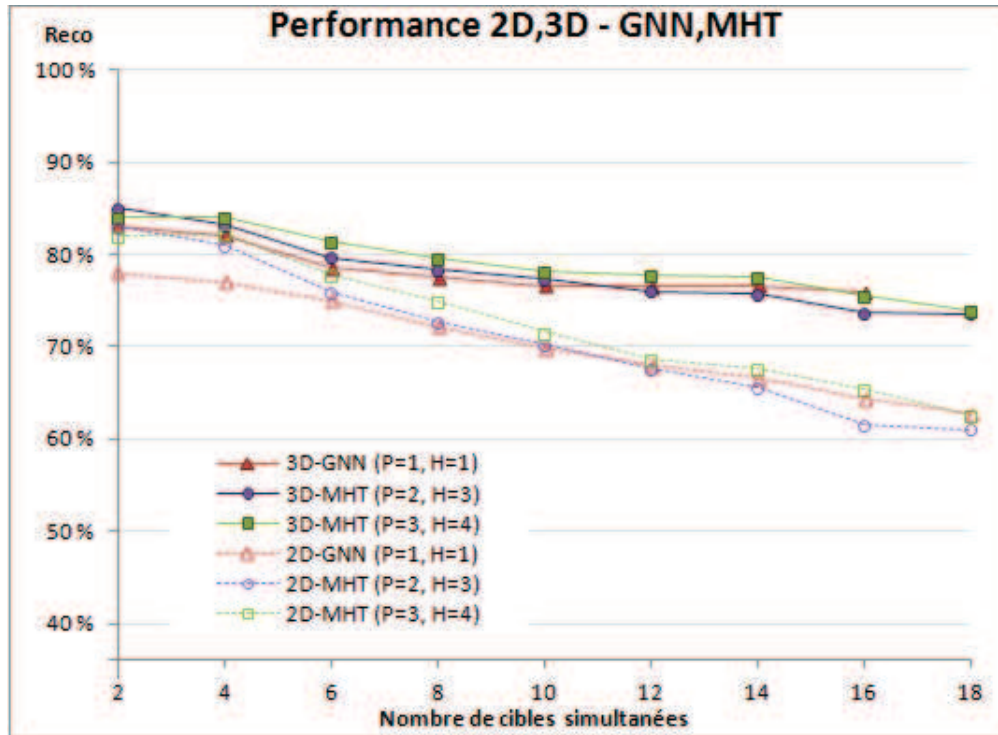


Figure 2.20 – Étude des performances 2D, 3D des approches GNN et MHT, où H est le nombre d'hypothèses et P le niveau d'élagage (ou "pruning"). Pour chaque nombre de cibles défini en abscisse, les résultats ont été calculés sur une centaine d'exemples différents instanciés à partir du modèle de scénario "Normal", soit un total 900 scénarios distincts.

Suite à ces 3 constats, nous retenons spécialement l'approche 3D-GNN pour la suite des expérimentations.

2.5.4.2 Importance de la méthode de re-projection pour la prise en compte des observations sans profondeur

MODÈLE DE SCÉNARIO UTILISÉ : "Normal"

Contrairement à la méthode de segmentation basique, notre segmentation HIDS détecte les cibles avec l'information de localisation 3D complète, mais détecte également les cibles sans information de profondeur. Ces dernières deviennent alors candidates à la re-projection. Dans cette expérience, nous avons tenté de mesurer l'intérêt de la prise en compte de ces cibles dont l'information est incomplète dans le suivi. Pour cela, nous avons retiré toutes les cibles sans information profondeur, ce qui a pour conséquence de retirer la majorité des fausses alarmes (soit environ 90% des cibles sans profondeur), mais également quelques vraies observations. Les résultats présentés dans la figure 2.21 ont confirmé la nécessité de prendre en compte toutes les observations (avec et sans profondeur) pour effectuer un suivi correct. Il est clair qu'en 3D, les difficultés sont liées aux observations sans profondeur ; néanmoins,

ces observations sont capitales pour assurer le suivi des traces concernées. Aussi, on constate qu'en 3D, la performance décroît finalement assez peu avec l'augmentation du nombre de cibles. En revanche, les erreurs de segmentation, quant à elles, ont une influence notable sur la performance.

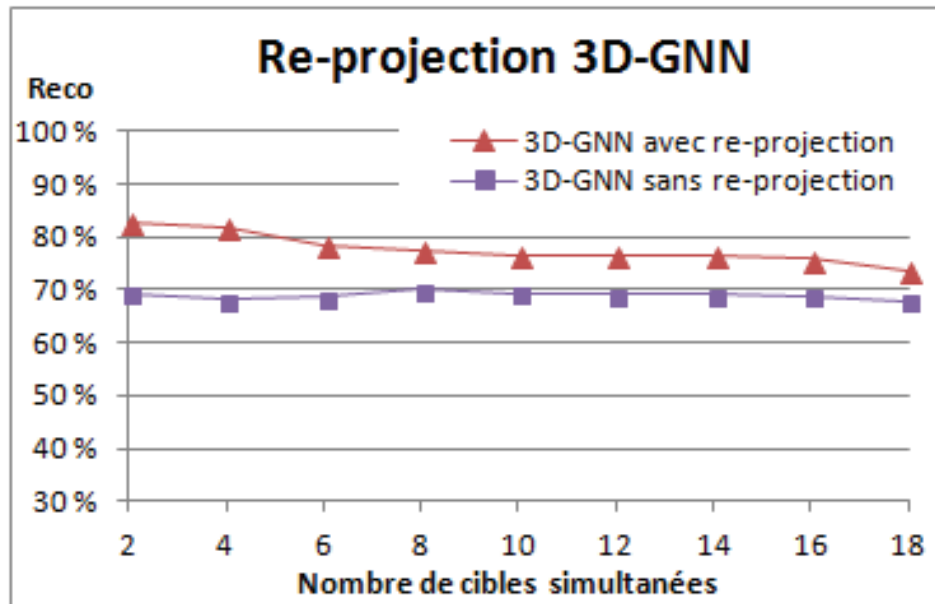


Figure 2.21 – Étude des performances de l'approche 3D-GNN, avec et sans, utilisation de la méthode de re-projection. Les expériences ont été réalisées sur $9 * 100 = 900$ scénarios distincts de type "Normal".

2.5.4.3 Étude de la robustesse de la méthode 3D-GNN avec différents taux de détections manquées (MD) et de fausses alarmes (FA)

MODÈLE DE SCÉNARIO UTILISÉ : "FA+" et "MD+"

On teste ici la méthode 3D-GNN avec, soit un taux doublé de MD, soit un taux doublé de FA. La figure 2.22 montre une meilleure robustesse face aux FA qu'aux MD. Cela confirme la stratégie adoptée au niveau de l'approche de détection, notamment sur le fait qu'elle renvoie des mesures ayant un rapport MD/FA faible.

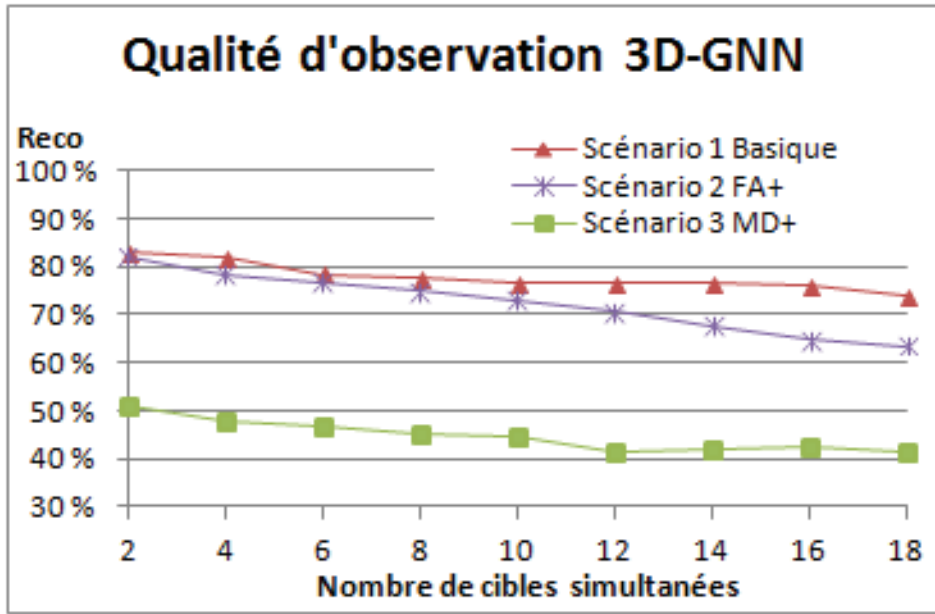


Figure 2.22 – Étude de la robustesse de l'approche 3D-GNN sous différentes conditions de segmentation. Les expériences ont été réalisées sur $3 \times 9 \times 100 = 2700$ scénarios distincts de type "Normal", "FA+" et "MD+".

2.5.4.4 Prise en compte du contexte de la scène :

MODÈLE DE SCÉNARIO UTILISÉ : "Couple"

Dans cette dernière partie, nous testons nos Contrib. 2a et 2b (voir les sections 2.3.2 et 2.3.3) permettant de prendre en compte le contexte environnemental de la scène pour améliorer le suivi. Soucieux d'effectuer une comparaison équitable, nous avons cherché pour chaque méthode étudiée, les jeux de paramètres optimaux (cf. figure 2.25).

La figure 2.24 montre que les 2 méthodes proposées améliorent les performances de suivi, en comparaison de la méthode classique. La Contrib. 2a fondée sur un *a priori* fort et contraignant offre le meilleur gain de performance. Quant à la Contrib. 2b, s'appuyant essentiellement sur les données, elle offre un gain de performance légèrement inférieur à la Contrib. 2a, mais a le mérite de s'affranchir du fort *a priori* nécessaire à cette dernière. Néanmoins, on note l'important temps de calcul (e.g. 10 sec/image) engendré par le parcours de la base de trajectoires dans la Contrib. 2b. Ceci nécessite un travail d'optimisation sur l'indexation des trajectoires. En guise d'illustration, la figure 2.23 présente les Contrib. 2a et 2b en action.

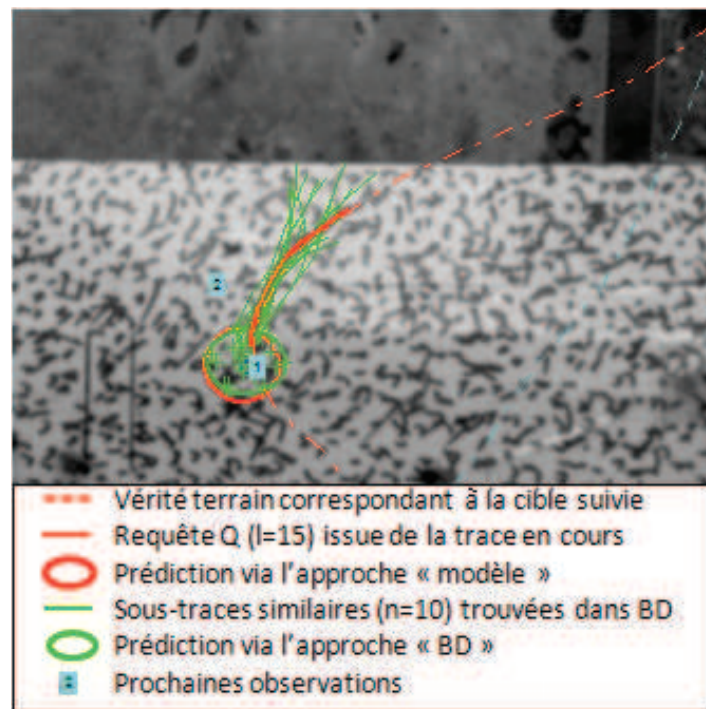


Figure 2.23 – Comparaison des 2 méthodes de suivi contraintes, l'une avec modèle (contrib. 2a) et l'autre sans modèle (contrib. 2b), sur un exemple d'abeille à l'approche de la planche d'envol. Les ellipses représentent les densités Gaussiennes estimées a priori.

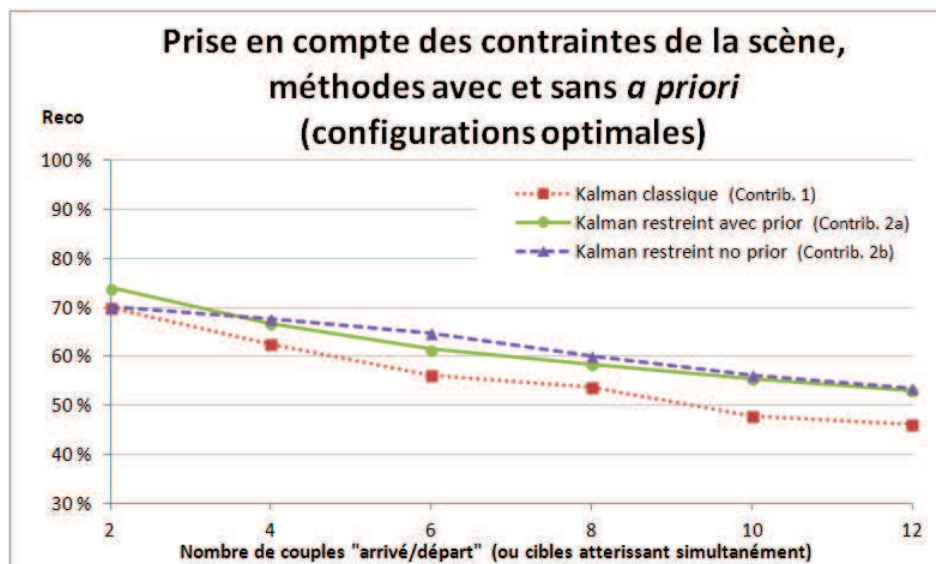


Figure 2.24 – Étude des performances apportées par nos Contrib. 2a et 2b (cf. Section 2.3) relativement à une approche classique. Les expériences ont été réalisées sur $6 \times 100 = 600$ scénarios distincts de type "Couple".

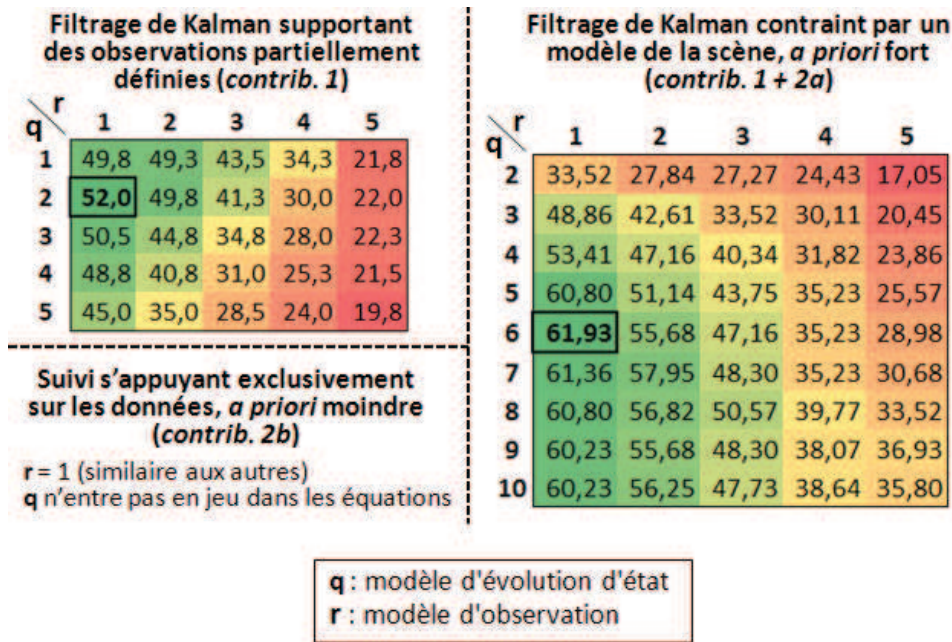


Figure 2.25 – Expériences pour le paramétrage empirique des modèles de bruit (q et r) du filtre de Kalman. Les chiffres indiquent les résultats moyens en fonction des différentes combinaisons de paramètres. Les résultats ont été calculés sur 25 scénarios de type 2 "Couple" (cf. Section 2.5.2), mettant en jeu 8 couples, 16 cibles. Les cases encadrées indiquent les jeux de paramètres optimaux. La solution proposée s'appuyant principalement sur les données ne fait pas intervenir le paramètre q (eq. 2.17).

2.6 Conclusion

Dans ce chapitre, nous nous sommes intéressés au problème de suivi en 3D de cibles à la dynamique chaotique à partir d'observations bruitées et/ou partielles. En premier lieu, nous avons constaté que les méthodes proposées dans la littérature ne permettaient pas de répondre à notre problématique de manière satisfaisante. Nous avons ainsi identifié 2 verrous scientifiques à lever :

- le problème des cibles localisées partiellement (information de profondeur manquante). Problème auquel nous avons répondu en proposant une première contribution. L'idée est de rendre les observations 2D exploitables dans un contexte de suivi 3D (Contrib. 1), s'appuyant temporairement sur une estimation de leur profondeur et ainsi permettre leur projection dans le repère de suivi. Les traitements sont ainsi effectués en coordonnées homogènes réelles (métrique), facilitant ainsi l'expression des modèles de dynamiques linéaires en 3D.

En effet, ces derniers seraient non-linéaires et complexes à appréhender dans l'espace image 2D (pixels) + carte de profondeur (métrique).

- l'incompatibilité apparente des contraintes "nombreuses cibles chaotiques" vs "agrégation auprès des zones d'intérêts" dans un environnement relativement peu contraint. Problème pour lequel nous avons proposé 2 approches sous la forme de contributions distinctes. L'une est fondée sur des hypothèses fortes sur le modèle de la scène et des

cibles (Contrib. 2a), tandis que l'autre par opposition s'appuie principalement sur les données observées, et est dénuée par conséquent d'*a priori* fort sur la dynamique et la forme des trajectoires (Contrib. 2b).

Les trois contributions proposées dans ce chapitre ont été évaluées à l'aide de scénarios simulés à partir d'une base de trajectoires réelles. La base de trajectoires est propre à notre étude. Les données (images brutes et/ou trajectoires) ont déjà été distribuées à plusieurs équipes de chercheurs des laboratoires L3i¹, MIA² et LSTA³.

Finalement, nous avons, à l'issue des deux premiers chapitres de cette thèse, proposé une méthodologie de détection et de suivi de cibles reposant sur une panoplie d'outils bien connus de la littérature, et proposé des améliorations/contributions les renforçant dans le cadre de contextes applicatifs 3D et issus de la mesures du vivant dans un milieu naturellement encombrés.

Dans le chapitre suivant, nous utilisons les trajectoires calculées grâce aux méthodes proposées pour faire de la modélisation comportementale.

1. Laboratoire L3i : Informatique, Image et Interaction, Avenue Michel Crépeau, 17042 La Rochelle Cedex 1

2. Laboratoire MIA : Mathématiques, Image et Applications, Avenue Michel Crépeau, 17042 La Rochelle Cedex 1

3. Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, 4 place Jussieu, 75005 PARIS

Modélisation comportementale par approche Bayésienne non-paramétrique

Sommaire

3.1	Introduction	86
3.2	État de l'art sur les modèles Bayésiennes non-paramétriques	87
3.2.1	Positionnement sur la modélisation comportementale	87
3.2.2	Présentation succincte des modèles Bayésiens non-paramétriques	88
3.2.3	Extensions du Processus de Dirichlet	92
3.3	Modèle MLC-HDP pour des trajectoires issues de séquences vidéo	97
3.3.1	Structure de données	97
3.3.2	Clustering selon trois niveaux	98
3.3.3	Signatures de trajectoires	100
3.3.4	Critère d'arrêt	102
3.4	Simulation Multi-Agents pour la validation de l'approche	103
3.4.1	Pertinence du couple SMA / MLC-HDP	103
3.4.2	Génération de la vérité terrain	103
3.4.3	Étude préliminaire	105
3.4.4	Étude approfondie	110
3.5	Conclusion	118

3.1 Introduction

Les travaux présentés dans les chapitres 1 et 2 ont permis de poser les bases pour la collecte en masse de trajectoires 3D à partir de scènes fortement encombrées. À présent, dans ce chapitre 3, nous étudions le problème de la modélisation comportementale d'un ensemble d'individus à partir de leurs trajectoires. Nous nous intéressons en particulier aux cas où les trajectoires sont de formes variées. Ici, l'aspect intéressant est la difficulté d'identifier des tendances ou de s'appuyer sur des hypothèses fortes pour établir un modèle de comportement. Nous avons été tout particulièrement confronté à ce problème dans le cadre de notre application apicole, qui est, rappelons le, le fil rouge de cette thèse. En fait, à partir des connaissances métiers apportées par les biologistes, nous avons rencontré des difficultés pour identifier des indices relatifs à la dynamique des abeilles en vol qui puissent nous être utile en termes de modélisation comportementale. Et c'est finalement cette difficulté qui a motivé les travaux présentés dans ce chapitre, soit une approche de modélisation ne nécessitant pas ou peu de connaissances métiers.

Problématique Nous posons alors la problématique suivante : "comment avec un minimum d'*a priori*, peut-on détecter et modéliser des comportements émergents à partir d'un ensemble structuré de trajectoires non contraintes". Cette problématique semble très large, mais derrière celle-ci se cache tout de même un cadre qui nécessite une structuration des données. Nous verrons par la suite que cette structure, bien que générique représente l'apport principal en termes de connaissance au sein de notre modèle de comportement.

Littérature Pour traiter cette problématique, nous nous orientons vers des approches Bayésiennes non-paramétriques, et en particulier vers des méthodes s'appuyant sur les processus hiérarchiques de Dirichlet (en anglais HDP pour Hierarchical Dirichlet Process). Depuis quelques années, malgré un contexte formel relativement complexe et des temps de calcul qui peuvent s'avérer importants, les modèles Bayésiens non-paramétriques connaissent un succès grandissant dans plusieurs domaines tels que la génétique [Beal 2006], la finance [Griffin 2006], et bien sûr, la vision par ordinateur. Pour ne citer que quelques exemples, on peut mentionner des applications liées à l'analyse de mouvements [Fox 2007], à la catégorisation de scènes [Fei-Fei 2005], à la reconnaissance d'objets [Sudderth 2005] ou à la reconnaissance d'activités humaines [Niebles 2008].

Approche proposée Dans le cadre de ce chapitre, nous mettons en œuvre une extension particulière du HDP, ayant la particularité de regrouper les données simultanément selon plusieurs niveaux obéissant à une structure hiérarchique. Nous verrons comment cette méthode peut être adaptée à notre problème de modélisation comportementale à partir de données de trajectoires.

Aujourd'hui les biologistes s'intéressent à des problèmes comme la modélisation du comportement d'essaims ou la modélisation de l'activité de colonies à partir de vidéos (e.g. [Schaefer 2012]). C'est d'ailleurs ce type d'applications qui a poussé la construction de notre chaîne de traitement et d'analyse de trajectoires. À notre connaissance, il n'existe pas à ce jour de vérité terrain faisant le lien entre trajectoires et comportements d'une colonie d'insectes. Pour palier ce manque, nous proposons, une méthode originale pour valider notre approche de modélisation comportementale. Celle-ci repose sur l'utilisation de Systèmes Multi-Agents

(SMA). Ces derniers permettent de définir des modèles de simulation de colonies d'insectes (e.g. fourmis, abeilles) capables à la fois de faire émerger des comportements et d'observer les trajectoires "non contraintes" associées à chacun des individus.

3.2 État de l'art sur les modèles Bayésiennes non-paramétriques et la modélisation comportementale

3.2.1 Positionnement dans le domaine de la modélisation comportementale

Dans la littérature, on trouve un large panel de travaux traitant du problème de la modélisation de comportements à partir de vidéos. Le lecteur peut se référer à l'étude [Morris 2008] qui récapitule un bon nombre d'approches existantes. On distingue les deux catégories suivantes :

Approches s'intéressant aux trajectoires individuelles Celles-ci exploitent des données de trajectoires obtenues à partir du suivi de cibles. Les trajectoires représentent de manière éparse l'essentiel de la dynamique de la vidéo. Les approches, s'appuyant sur les trajectoires, s'intéressent le plus souvent aux signatures invariantes et se focalisent principalement sur l'information spatiale locale. Pour ne citer que quelques exemples, dans [Stauffer 2000], les trajectoires sont regroupées en activités à partir, entre autres, des positions, vitesses et directions. Suivant le même principe, dans [Wang 2006] est utilisée une distance de Hausdorff modifiée pour comparer les trajectoires entre elles. On retrouve dans d'autres travaux une méthode de découpage des trajectoires en primitives [Hongeng 2001, Feldman 2004]. Celles-ci sont souvent obtenues à l'aide d'un classifieur entraîné sur une base de données étiquetées. On trouve aussi des travaux qui proposent de modéliser les interactions entre les cibles, par exemple avec un Modèle de Markov Caché (ou HMM pour Hidden Markov Model) [Oliver 2000] ou des champs aléatoires de Markov (ou MRF pour Markov Random Field) [Khan 2005]. Dans [Intille 1999], un réseau bayésien est utilisé pour analyser les stratégies d'un jeu de football, mais le suivi de cibles automatisé n'ayant pas fonctionné, les trajectoires des joueurs ont été annotées à la main.

Approches s'appuyant sur l'estimation du mouvement Celles-ci exploitent des données de mouvement telles que des vecteurs calculés par une méthode de flot optique. Ces vecteurs de mouvement représentent, pour chaque image de la vidéo, la dynamique de la scène dans son ensemble de manière dense. Ces approches sont plutôt utilisées sur des scènes encombrées pour lesquelles un suivi de cibles de manière individualisée n'est pas envisageable. Ces approches s'appuient sur l'évolution du mouvement et attribuent généralement plus d'importance à la dimension temporelle que spatiale pour la modélisation des comportements. Se pose alors le problème de savoir quand un comportement (ou une activité) commence et se termine. Ces problèmes sont abordés dans [Zelnik-Manor 2001] (modélisation et clustering de clips vidéo à l'aide d'un histogramme multirésolution), ou [Zhong 2004] (analyse de type sujets/documents sur de la vidéo). Mais il reste à résoudre le problème de la modélisation des activités se déroulant simultanément. En réponse, des travaux proposent de découper le concept d'activité en différentes actions [Rao 2002, Gorelick 2007].

Comme nous l'avons mentionné dans l'introduction, ces dernières années ont montré un intérêt grandissant pour les modèles Bayésiens non-paramétriques, notamment dans le do-

maine de la modélisation d'activités ou de comportements. Les modèles Bayésiens hiérarchiques fournissent d'ailleurs une solution attractive pour la modélisation à partir de données chaotiques. Le HDP originalement proposé dans [Teh 2006], est une approche Bayésienne non-paramétrique permettant de clusteriser un jeu de données à différentes échelles, tout en partageant l'information au travers d'une hiérarchie. Des approches similaires ont été largement utilisées par la communauté document (e.g. LDA, pour Latent Dirichlet Allocation, dans le cadre de l'analyse de sujets/documents). Récemment, ces approches HDP ont montré leur efficacité pour regrouper des données de natures différentes. Concernant la modélisation d'activités à partir de vidéos, la majorité des travaux s'appuyant sur des modèles Bayésiens non-paramétriques, quelque soit la structure de données sous-jacente, utilise des données de mouvement issues du flot optique, où les vidéos sont scindées en entités temporelles (e.g. quelques secondes) [Kuettel 2010, Wang 2009, Hospedales 2009]. Il existe relativement peu de travaux reposant sur des modèles Bayésiens non-paramétriques à partir de données de trajectoires. Dans le domaine médical, on peut citer [Wang 2011] qui modélise des tractographies (données proches des trajectoires) à l'aide d'un modèle de mélange de HDPs.

Dans la suite de l'état de l'art, nous nous intéresserons aux modèles Bayésiens non-paramétriques, et verrons notamment comment ces derniers peuvent être utilisés pour la modélisation de comportements à partir de trajectoires.

3.2.2 Présentation succincte des modèles Bayésiens non-paramétriques

Cette section a pour objectif d'accompagner le lecteur en apportant les principes nécessaires à la compréhension de ce qu'est le clustering de Dirichlet. Néanmoins, cette section n'a pas pour ambition de dérouler un état de l'art exhaustif sur les méthodes Bayésiennes non-paramétriques, ce qui sortirait du cadre de cette thèse. Pour plus de détails sur les fondements, on conseille au lecteur de se reporter aux tutoriaux de référence [Teh 2007, Teh 2010, Orbanz 2011, Walker 1999]. Ceux pour qui ces notions sont familières, peuvent passer directement à la section 3.2.3 où sont abordées certaines extensions du processus de Dirichlet telles que le MLC-HDP (pour Multi Layer Clustering HDP) [Wulsin 2012] qui est le maillon clé de notre approche.

3.2.2.1 Loi binomiale / Loi beta

La loi binomial de paramètres n et p est une loi de probabilité discrète qui correspond à l'expérience suivante : "le renouvellement n fois, de manière indépendante, d'une épreuve de Bernoulli de paramètre p ". L'épreuve de Bernoulli est une expérience aléatoire à deux issues possibles : "succès" avec la probabilité p , ou sinon "échec" avec la probabilité $1 - p$. Finalement, on note k est le nombre de succès à l'issue des n tirages. La fonction de densité de probabilité de la loi binomiale est notée :

$$f(k|n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (3.1)$$

Maintenant, retournons le problème. On cherche à estimer la densité de probabilité du paramètre p à partir des n observations. Tout d'abord, rappelons qu'une loi conjuguée modélise la densité de probabilité du paramètre d'une loi sachant sa réalisation. Dans notre cas, la loi Binomiale a pour conjuguée la loi Beta, avec respectivement $\alpha - 1$ et $\beta - 1$ les nombres de

succès et d'échecs observés. Ainsi, la fonction de densité de probabilité de p sachant α et β est notée :

$$f(p|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (3.2)$$

avec B la fonction Beta notée :

$$B(x, y) = \frac{(x-1)!(y-1)!}{(x+y-1)!} \quad (3.3)$$

La figure 3.1 illustre les fonctions de densités de la loi binomiale et de sa conjuguée. Sur cet exemple, la loi binomiale de paramètres $p=0.3$ et $n=30$ donne le maximum de la densité de probabilité pour $k=9$ succès. De manière complémentaire, la loi Beta de paramètres $\alpha-1=9$, et donc $\beta-1 = n-9 = 21$, donne un maximum pour $p=0.3$.

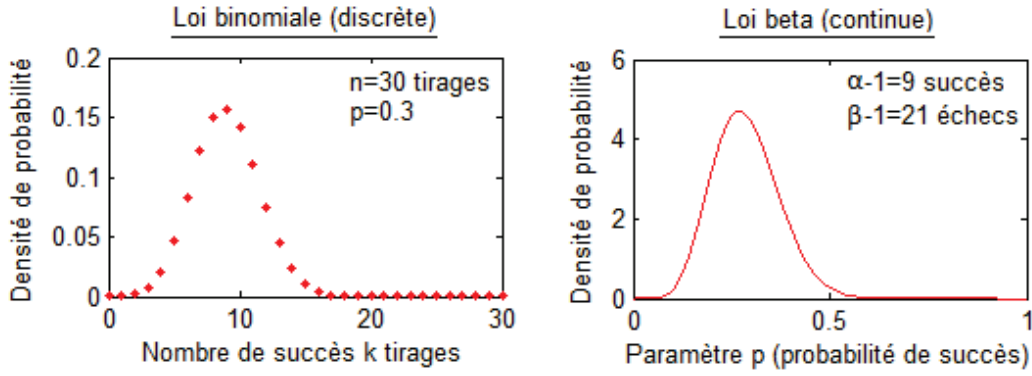


Figure 3.1 – Exemple de fonctions de densité de probabilité de la loi de Binomiale et de sa conjuguée, la loi Beta.

3.2.2.2 Loi multinomiale / Loi de Dirichlet

La loi binomiale ne prend en compte que deux issues possibles, ce qui est souvent illustré par l'exemple de la pièce de monnaie tombant sur "pile" ou "face". Maintenant, étendons le principe à la loi multinomiale. Celle-ci va permettre de prendre en compte plusieurs résultats. Elle peut être illustrée par n jets d'un dé à K faces, où chaque face a une probabilité p_k de succès. La somme des probabilités p_k est égale à 1. La densité de la loi multinomiale est notée :

$$f(x|n, p) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \quad (3.4)$$

où $x = (x_1, \dots, x_k)$ est le nombre de succès de chaque face pour n tirages suivant les probabilités $p = (p_1, \dots, p_k)$ relatives à chaque face du dé. Maintenant, retournons le problème une nouvelle fois tel que nous l'avons fait pour la loi binomiale. On cherche alors à estimer la densité de probabilité des p_k à partir des n observations. Pour cela, on utilise alors la loi de Dirichlet (généralisation de la loi Beta) qui est la conjuguée de la loi multinomiale, où $\alpha = (\alpha_1, \dots, \alpha_k)$ représente le nombre d'observations de chaque face. La densité de probabilité de la loi de Dirichlet est notée :

$$f(p_1, \dots, p_k | \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^k p_i^{\alpha_i - 1} \quad (3.5)$$

avec p_1, \dots, p_k les paramètres à estimer. On note la constante de normalisation B exprimée à l'aide de la fonction Gamma.

$$B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)} \quad \text{et} \quad \Gamma(n) = (n-1)! \quad (3.6)$$

Par la suite, on notera la loi de Dirichlet $\text{Dir}(\alpha_1, \dots, \alpha_K)$. La figure 3.2 illustre les fonctions de densité de la loi multinomiale et de sa conjuguée sur un exemple de dé à 3 faces. La loi multinomiale de paramètre $p = (0.5, 0.2, 0.3)$ pour $n = 10$ tirages, donne le maximum de densité de probabilité pour $x = (5, 2, 3)$ succès. De manière complémentaire, la loi de Dirichlet de paramètre $\alpha = (5, 2, 3) + 1$ donne le maximum de densité de probabilité pour $p = (0.5, 0.2, 0.3)$. L'idée est que, plus le nombre de tirages augmente, plus la densité vient se concentrer autour de ce maximum.

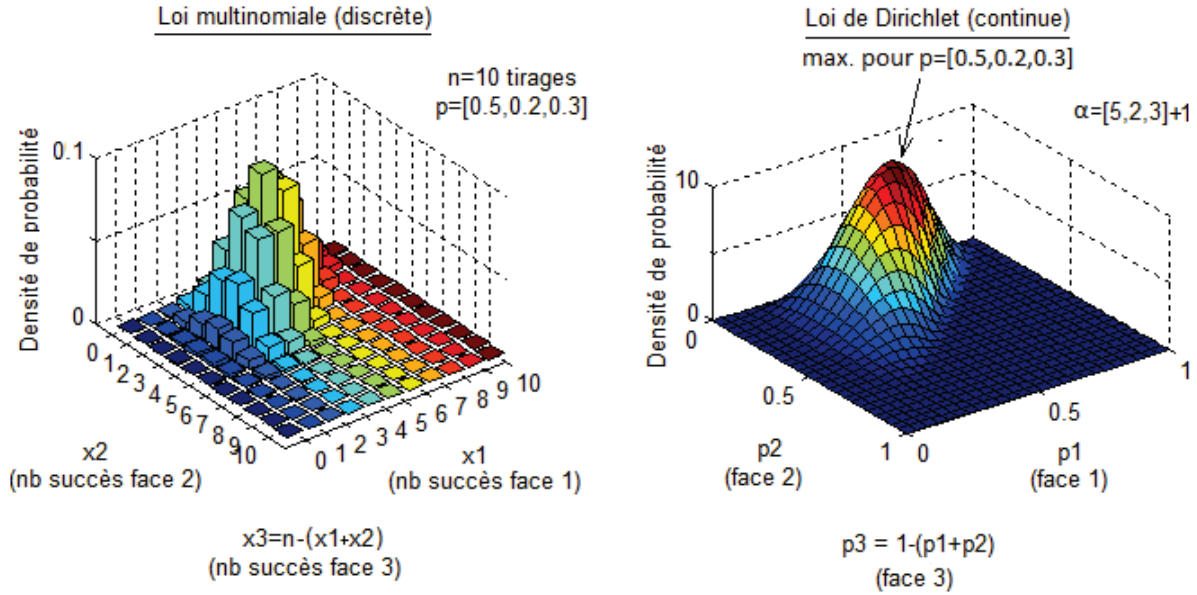


Figure 3.2 – Exemple de fonctions de densité de probabilité : loi multinomiale et sa conjuguée, la loi de Dirichlet.

3.2.2.3 Processus de Dirichlet (DP) et construction "stick-breacking"

Le processus de Dirichlet (DP) [Ferguson 1973] est une généralisation de la loi de Dirichlet pour le cas où k est infini. Un DP permet de tirer une distribution presque sûrement discrète (appelée G) à partir d'une distribution de probabilité continue (appelée H ou G_0). Le tirage de G par un DP est noté $G \sim DP(\alpha, H)$, avec α (parfois noté γ) un scalaire représentant la concentration (ou paramètre d'échelle). On note pour toutes partitions A_1, \dots, A_k d'un espace mesurable :

$$(G(A_1), \dots, G(A_k)) \sim \text{Dir}(\alpha H(A_1), \dots, \alpha H(A_k)) \quad (3.7)$$

La figure 3.3 illustre la notion de partitionnement infini.

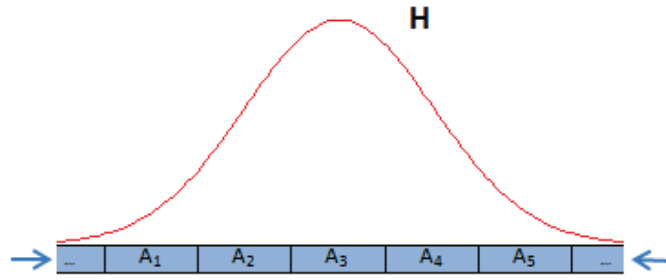


Figure 3.3 – Exemple où H est une loi de probabilité normale (1 dimension). La densité est partitionnée, et lorsque $k \rightarrow \infty$, chaque A_i devient une impulsion de Dirac.

Les DP peuvent être construits de plusieurs manières : la représentation du "bâton cassé" (ou en anglais "stick-breaking") [Sethuraman 1994], la métaphore du restaurant chinois [Teh 2006] ou également la méthode des urnes de polya [Blackwell 1973]. Nous détaillons ci-dessous la construction stick-breacking.

L'objectif est d'associer des poids π (ou masses) tendant à être de plus en plus faibles à l'infinité des impulsions de Dirac (appelées atomes) support de G . Rappelons que ces atomes sont tirés via la loi de Dirichlet (conjuguée de la loi multinomiale) définie par H . Le paramètre α contrôle la distribution des masses à travers les atomes. Plus α est grand, plus la masse est concentrée sur les premiers atomes. La notation suivante accompagnée de la figure 3.4 illustre le principe :

$$\begin{aligned} \beta_i &\sim \text{Beta}(1, \alpha) \\ \pi_i &= \beta_i \prod_{l=1}^{i-1} (1 - \beta_l) \\ \phi_i &\sim H \\ G &= \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k} \end{aligned} \quad (3.8)$$

Cette mesure est parfois notée $\pi \sim GEM(\alpha)$ faisant référence à Griffiths, Engen, and McCloskey [Pitman 2002].

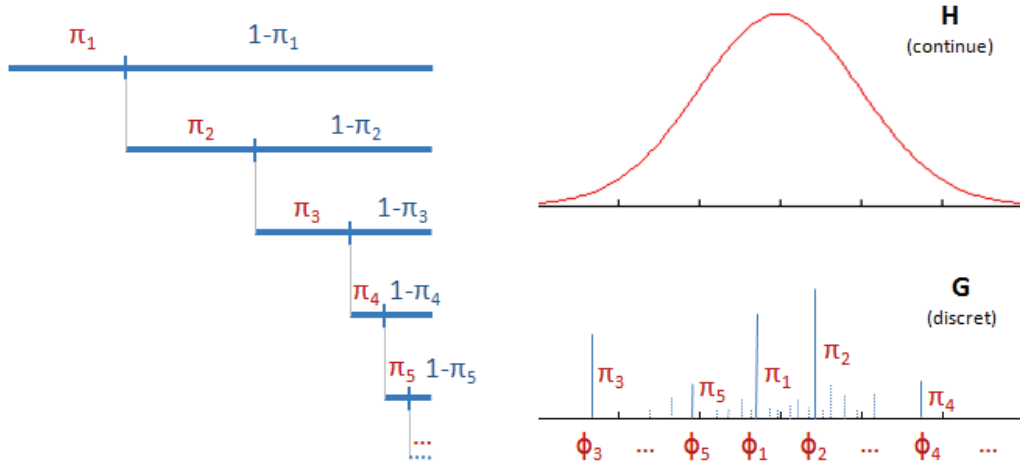


Figure 3.4 – Construction stick-breacking du support de la distribution discrète G à partir de la distribution H .

3.2.3 Extensions du Processus de Dirichlet

3.2.3.1 Processus de Dirichlet à mélange (DPM)

Grâce aux Processus de Dirichlet à Mélange (DPM ou Mixtures par Processus de Dirichlet), il devient possible de faire du clustering sur des données avec un nombre de clusters k théoriquement infini. L'aspect clustering vient du fait qu'à partir d'un DP, on peut tirer une distribution discrète G constituée d'atomes où chacun représente un cluster potentiel. Considérant un tirage à partir de G , les chances de retomber sur les mêmes clusters de poids élevé provoquent un effet clustering.

De manière formelle, on suppose qu'une observation x_i est tirée d'une distribution F (par exemple une gaussienne) de paramètre θ_i , lui même tiré de G , issue d'un DP. Ci-dessous la notation d'un DPM, accompagnée de la figure 3.5 pour sa représentation graphique :

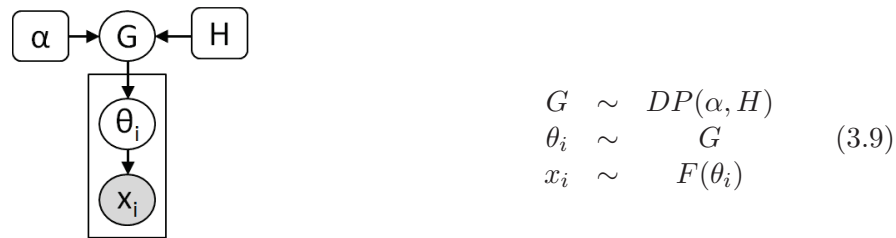


Figure 3.5 – Représentation graphique et équations d'un DPM.

Notons que la loi *a posteriori* d'un DP après n observations $(\theta_1, \dots, \theta_n)$ est aussi un processus de Dirichlet, noté :

$$P(G|\theta_1, \dots, \theta_n) = DP\left(\frac{\alpha}{\alpha + n}H + \frac{1}{\alpha + n} \sum_{i=1}^n \delta_{\theta_i}, \alpha + n\right) \quad (3.10)$$

Cela permet notamment d'estimer H , la distribution de base sur l'espace des paramètres

(e.g. μ et θ de lois gaussiennes) servant à modéliser les clusters à partir des observations. Après plusieurs répétitions au sein d'un processus itératif, on conserve la solution qui offre la plus grande vraisemblance. Finalement, l'algorithme pour la mise en pratique est relativement simple (cf. Algorithme 2).

```

while Observation Iteration do
  for each point X in Data do
    1) Remove  $X$  from its assigned cluster.
    2) Re-Estimate parameters for each cluster (without  $X$ )
    3) Compute the probability vector (length of active clusters) being the
       likelihood to be generated by each cluster, with  $\alpha$  at the last position (opening
       the opportunity for new cluster creation)
    4) Sample the new class of  $X$  according to the Multinomial distribution given
       by the DP.
    5) If last position is chosen, a new cluster is set to be active (should be sampled
       from  $G$ ), but in practice, we can estimate the parameters from  $X$ 
  end
end

```

Algorithme 2 : Algorithme pour la mise en pratique du clustering par DP.

La figure 3.6 montre un exemple de clustering sur des données aléatoirement générées par un mélange de gaussiennes. On note la capacité du clustering par DP à proposer des solutions avec des clusters se chevauchant (ou se superposant complètement), contrairement à une approche classique de type K-means.

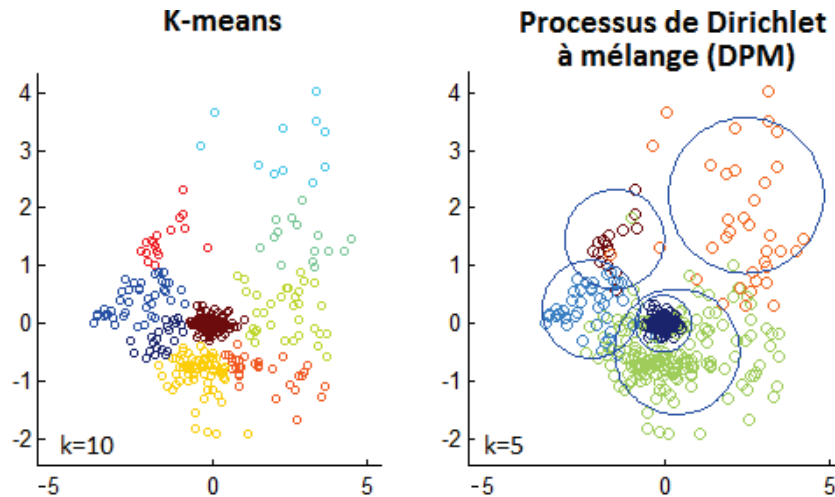


Figure 3.6 – Comparaison de la méthode K-means et DP sur des données générées aléatoirement à partir d'un mélange de gaussiennes. À gauche, pour la méthode K-means, K est calculé automatiquement via une méthode agglomérative à base de dendrogrammes. À droite, le résultat correspond à la 100ème itération d'un DP arrivé à un niveau relativement stable.

3.2.3.2 Processus Hiérarchique de Dirichlet (HDP)

Le Processus Hiérarchique de Dirichlet (ou HDP pour Hierarchical Dirichlet Process) [Teh 2006] est une manière de clusteriser des données potentiellement divisées en plusieurs groupes. La classification non-supervisée s'effectue au sein de chaque groupe indépendamment. Les clusters sont cependant partagés entre les groupes (cf. Figure 3.8). Tout comme un DP, le HDP permet de trouver le nombre de clusters et leurs paramètres susceptibles d'expliquer les données en s'appuyant sur les données elles-mêmes. Mais le principal intérêt du HDP réside dans sa capacité à partager l'ensemble des clusters au sein des différents groupes. À chaque cluster est associé un poids (possiblement nul) par groupe. Ainsi, les clusters peuvent n'être associés à aucune donnée dans certains groupes. Un HDP est défini par un DP classique, et plusieurs sous-DPs dépendants (1 associé à chaque groupe). Le DP classique offre le support sur lequel les sous-DPs vont s'appuyer pour échantillonner des observations. Soit la notation suivante, accompagnée de sa représentation graphique :

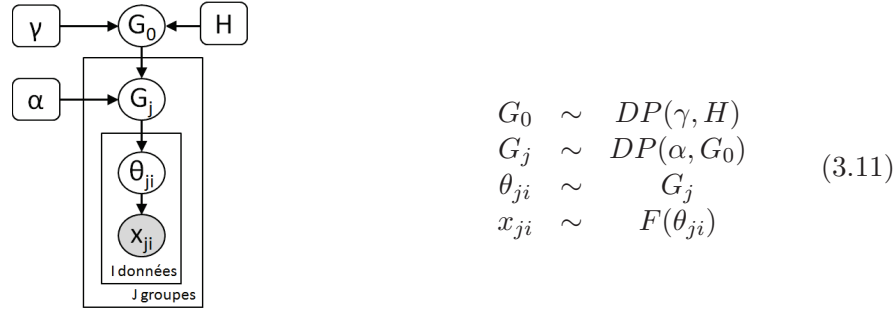


Figure 3.7 – Représentation graphique et équations d'un HDP.

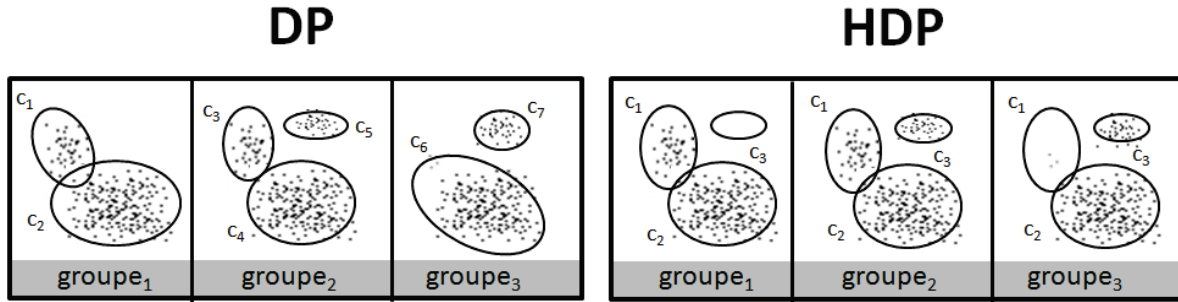


Figure 3.8 – Illustration de la faculté de partage des clusters de la méthode HDP sur des données structurées en groupes, comparée à l'approche DP classique.

La nouvelle construction "stick-breaking" pour le HDP est donnée par :

$$\begin{aligned}
 G_0 &= \sum_{k=1}^{\infty} \pi_{0k} \delta_{\theta_k^*} & \theta_k^* &\sim H \\
 \pi_{0k} &= \beta_{0k} \prod_{l=1}^{k-1} (1 - \beta_{0l}) & \beta_{0k} &\sim \text{Beta}(1, \gamma) \\
 G_j &= \sum_{k=1}^{\infty} \pi_{jk} \delta_{\theta_k^*} & & \\
 \pi_{jk} &= \beta_{jk} \prod_{l=1}^{k-1} (1 - \beta_{jl}) & \beta_{jk} &\sim \text{Beta}(\alpha \beta_{0k}, \alpha(1 - \sum_{l=1}^k \beta_{0l}))
 \end{aligned} \quad (3.12)$$

3.2.3.3 Autres variantes (NDP, DDP...)

Parmi les variantes notables, on peut mentionner le Nested HDP. En fait, avec le HDP, on considère que l'on connaît la structure en groupe des données observées. En pratique, il peut s'avérer que cette structure définie préalablement ne reflète pas vraiment la réalité en termes de regroupement des données. Il se peut que certains groupes soit plus cohérents que d'autres. C'est dans ce cadre que [Rodríguez 2008] propose le "Nested HDP" permettant de clusteriser les groupes grâce à un second niveau de mélange de modèles.

Autre variante, le Dependent Dirichlet Process introduit par [MacEachern 1999] permet de modéliser l'évolution progressive des clusters dans l'espace et dans le temps.

Sinon, on peut mentionner le HDP-HMM [Fox 2008] qui permet de découvrir les états types d'un modèle de Markov caché ainsi que les probabilités de transition associées.

3.2.3.4 Processus hiérarchique de Dirichlet multi-niveaux (MLC-HDP)

Le clustering par processus hiérarchique de Dirichlet multi-niveaux introduit par [Wulsin 2012] combine les avantages du HDP et du NDP. Il intègre l'aspect clustering à plusieurs niveaux comme pour le NDP tout en gardant la capacité du HDP à fournir un support commun pour le partage des clusters au travers des données. La flexibilité du MLC-HDP permet de définir le nombre de niveaux en fonction de la structure de données utilisée. Voici par exemple la définition d'un MLC-HDP à deux niveaux, et sa représentation graphique associée.

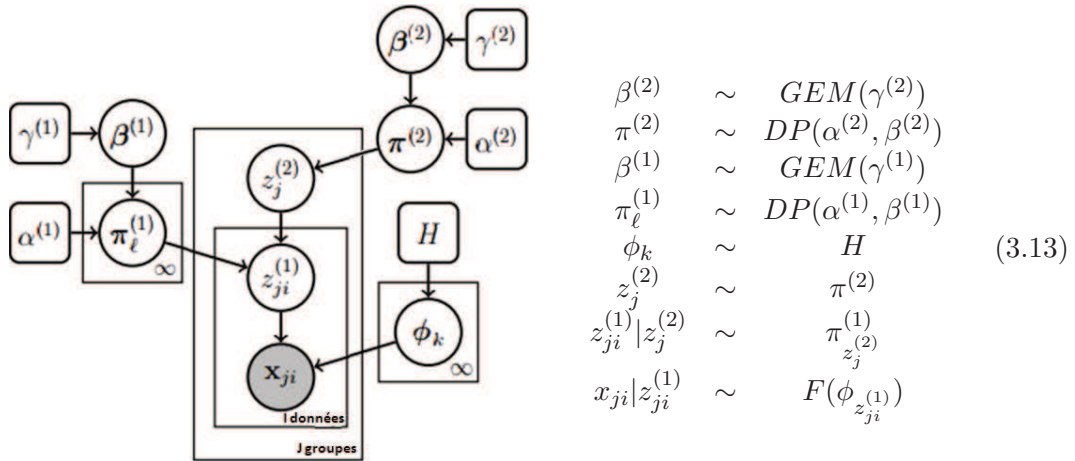


Figure 3.9 – Représentation graphique et équations d'un MLC-HDP à deux niveaux.

Dans le modèle MLC-HDP défini ci-dessus, les indices supérieurs indiquent le niveau au sein de la structure hiérarchique. $\pi^{(n)}$ représente l'ensemble des poids associés à chaque cluster au niveau n . L'indice inférieur associé à π renseigne le cluster de niveau supérieur décrit. z indique le cluster d'appartenance d'une donnée (niv. = 1) ou bien d'un cluster (niv. > 1). Le modèle s'appuie sur un certain nombre d'hyper-paramètres (α et γ à chaque niveau), néanmoins, l'auteur a montré que l'*a priori* porté par ces hyper-paramètres a une moindre influence sur les résultats finaux [Wulsin 2012].

3.2.3.5 Bilan DP, HDP, NDP, MLC-HDP

De manière anticipée, l'exemple ci-après utilise des trajectoires, ce qui permettra de faire le lien avec la section suivante. La figure 3.10 illustre donc à partir de trajectoires le fonctionnement des 4 principaux modèles présentés dans l'état de l'art de ce chapitre, à savoir le DP, le HDP, le NDP et le MLC-HDP. Notons que le DP se contente de regrouper les trajectoires similaires en clusters. L'ensemble des clusters est représenté par une mesure discrète G sur l'espace des paramètres θ (e.g. μ et σ modélisant chaque cluster par une Gaussienne). Le HDP donne lieu à une mesure discrète G_0 et des mesures G_i qui partagent les atomes avec G_0 mais associées à des poids différents. Le NDP donne lieu à une infinité de mesures discrètes G_i^* contenant chacune une infinité d'atomes et de poids. Le MLC-HDP donne également lieu à une potentielle infinité de mesures discrètes $G_i^{(1)}$, mais ce sont des échantillons de DP's issus d'une mesure de base $G_0^{(1)}$.

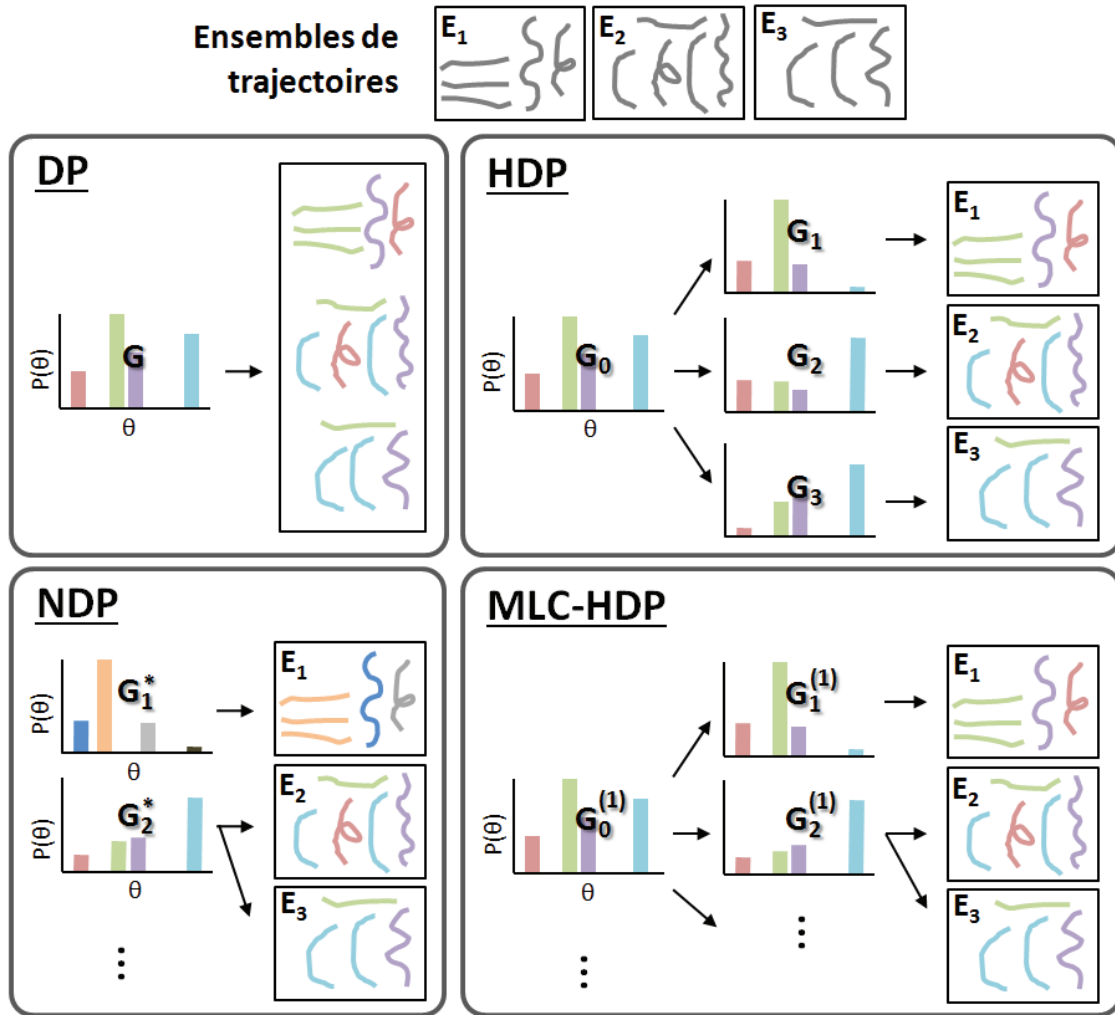


Figure 3.10 – Comparaison des 4 modèles : Processus de Dirichlet (DP), le processus hiérarchique de Dirichlet (HDP), le processus de Dirichlet imbriqués (NDP) et le clustering HDP multi-niveaux (MLC-HDP). L'exemple utilisé est un jeu de données constitué de 3 ensembles de trajectoires. Chaque couleur correspond à un cluster (figure adaptée de [Wulsin 2012]).

3.3 Modèle MLC-HDP pour des trajectoires issues de séquences vidéo

Les approches Bayésiennes non-paramétriques de type LDA ou HDP ont récemment montré leur efficacité en termes de modélisation comportementale de scènes encombrées. Nous proposons ainsi dans cette section d'utiliser le MLC-HDP [Wulsin 2012] (originellement proposé dans le cadre d'une application médicale), pour modéliser des comportements à partir de données structurées de trajectoires. Le MLC-HDP a l'avantage de gérer des structures de données possédant plus de deux niveaux hiérarchiques, ce qui, à notre connaissance, va au delà des approches similaires existantes [Kuettel 2010, Wang 2009, Hospedales 2009]. De plus, plutôt que s'appuyer, comme il se fait classiquement pour les scènes encombrées, sur des données de mouvement dense (e.g. flot optique), nous proposons d'utiliser des données éparées, et en particulier des trajectoires individuelles.

Notre objectif est d'établir les propriétés émergentes d'une dynamique d'agents à partir de l'expression, sous forme de signatures, des comportements individuels. Ces dynamiques mobilisent des processus opérant à des échelles de temps différentes. Pour l'atteindre, nous choisissons d'établir cette réduction à travers une classification (agrégation) sur plusieurs niveaux de description (DP hiérarchique) :

- les agents obéissent à des dynamiques de groupes (i.e. liés à leur rôle dans la ruche). Les méthodes de clustering/classification permettent d'agréger ces différentes dynamiques au sein de clusters ;
- si le cluster est suffisamment représentatif de la dynamique pour un groupe donné, c'est qu'une certaine distance entre les observations et ce modèle reste inférieure à une certaine valeur sur une plage de temps significative. C'est donc l'ensemble des clusters, et leurs paramètres, qui sont capturés par le modèle agrégé ;
- pour prendre en compte les différentes échelles d'observation, nous proposons une découpage temporel hiérarchique des observations.

Nous supposons que la production d'un modèle agrégé permettra, (1) de capturer ce qui est essentiel dans les dynamiques observées et (2) d'exprimer des aspects qualitatifs et quantitatifs importants. In fine, il doit permettre de relier les dynamiques observées aux valeurs des paramètres comportementaux comme nous le montrerons par la suite à l'aide de simulations multi-agents.

3.3.1 Structure de données

Le jeu de données utilisé dans notre approche est structuré sous la forme d'un ensemble de "*captures*"¹. Les *captures* correspondent à une même scène filmée par une caméra fixe à différents instants (e.g. heures, jours, semaines). Les *captures* sont segmentées en "*sousclips*" à l'aide d'une fenêtre glissante non-chevauchante (e.g. 5 sec). La taille de la fenêtre doit être suffisamment grande pour assurer un nombre suffisant de trajectoires dans chaque *sousclip*, mais doit aussi être de taille raisonnable pour éviter que les *sousclips* contiennent plusieurs

1. Afin de faciliter la lecture, les différents concepts introduits entre guillemets dans cette section seront repris en italique dans la suite du manuscrit.

"activités". Le concept d'*activité* sera défini par la suite à l'aide d'exemples. Finalement, le jeu de données est constitué de nombreuses trajectoires structurées en *captures* et *sousclips* suivant une granularité donnée. La figure 3.11 illustre notre structure de données selon 3 niveaux et la compare avec celle utilisée par D. Wulsin dans son application de modélisation de crises d'épilepsie.

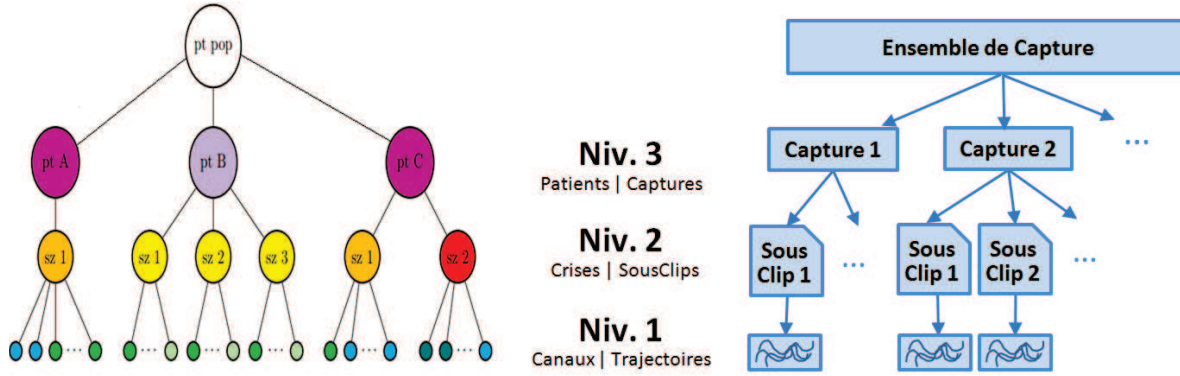


Figure 3.11 – Adaptation de la structure de données utilisée dans [Wulsin 2012] "Patients, Crises d'épilepsie, Canaux" (à gauche) pour notre application structurée en "*captures, sousclips, trajectoires*" (à droite).

3.3.2 Clustering selon trois niveaux

Allant de pair avec la structure de données, nous utilisons une version à trois niveaux du MLC-HDP, où les éléments (*captures, sousclips, trajectoires*) issus d'un même niveau sont regroupés au sein de clusters types en fonction de leurs similarités. Voici comment le modèle à trois niveaux s'intègre avec les données.

Niveau 1 : les trajectoires, ou plutôt signatures de trajectoires sont regroupées en clusters appelés "*atomes d'activité*". Par exemple, un cluster peut regrouper des cibles (ou plutôt des trajectoires) se dirigeant dans la même direction ;

Niveau 2 : chaque *sousclip* est modélisé par une expression pondérée de tous les *atomes d'activité*. Les modèles d'expressions (ou ensembles de poids) d'*atomes d'activité* proches sont regroupés en modèles types que l'on appelle *activités*. Donc, à chaque *sousclip* est associée une *activité*. Par exemple, une *activité* peut modéliser le fait que les individus d'une colonie effectuent des allers/retours entre un point A et un point B ;

Niveau 3 : chaque *capture* est modélisée comme une expression pondérée de toutes les *activités*. Les modèles d'expressions (ou ensembles de poids) d'*activité* proches sont regroupés en modèles types que l'on appelle "*comportements*". Donc, à chaque *capture* est associée un *comportement*. Par exemple, un *comportement* peut modéliser le fait qu'une colonie d'individus

passé 20% du temps à explorer l'espace, et 80% du temps à exploiter une ressource de manière efficace.

De manière formelle, nous considérons un jeu de données constitué de T *captures*, où chaque *capture* t est composée de J_t *sousclips*, où chaque *sousclip* j est composé de N_{tj} signatures de trajectoires, et où chaque signature i s'écrit $x_{tji} \in R^d$, avec d la dimension spatiale des signatures. L'observation $\{\{\{x_{tji}\}_{i=1}^{N_{tj}}\}_{j=1}^{J_t}\}_{t=1}^T$ de toutes les trajectoires est modélisée par un ensemble de processus hiérarchiques de Dirichlet. Sur le niveau le plus bas (niv. 1), les trajectoires sont regroupées au sein de clusters définis par des DPs dont les paramètres sont tirés d'une distribution génératrice H . Les paramètres γ et α sont les hyperparamètres du modèle. Nous considérons F comme une distribution normale multivariée avec une matrice de covariance diagonale. Les trois niveaux du MLC-HDP sont définis par :

Niv.3 - <i>Comportement</i>	Niv.2 - <i>Activité</i>	Niv.1 - <i>Atome d'activité</i>
1 Processus de Dirichlet agrégeant les <i>captures</i> en <i>comportements</i> similaires.	T DPs (1 par <i>capture</i>) agrégeant les <i>sousclips</i> en <i>activités</i> similaires.	J_t DPMs (1 par <i>sousclip</i>) agrégeant les trajectoires similaires en cluster selon un modèle de mixture de gaussiennes.
$\beta^{(3)} \sim GEM(\gamma^{(3)})$ $\pi^{(3)} \sim DP(\alpha^{(3)}, \beta^{(3)})$ $z_t^{(3)} \sim \pi^{(3)}$	$\beta^{(2)} \sim GEM(\gamma^{(2)})$ $\pi_l^{(2)} \sim DP(\alpha^{(2)}, \beta^{(2)})$ $z_{tj}^{(2)} \sim \pi_{l=z_t^{(3)}}^{(2)}$	$\beta^{(1)} \sim GEM(\gamma^{(1)})$ $\pi_\ell^{(1)} \sim DP(\alpha^{(1)}, \beta^{(1)})$ $z_{tji}^{(1)} \sim \pi_{\ell=z_{tj}^{(2)}}^{(1)}$ Trajectoires (signatures) $\phi_k \sim H$ $x_{tji} \sim F(\phi_{k=z_{tji}^{(1)}})$

où le paramètre d'intensité α régule le nombre de pics. Les données x_{tij} (signatures de trajectoire) sont tirées d'une distribution F paramétrée par ϕ_k . Les paramètres ϕ_k sont eux mêmes issus d'une distribution génératrice H . Afin de clarifier le lien entre le modèle et le jeu de données, la figure 3.12 montre comment le jeu de données hiérarchiquement structuré est lié aux différentes couches du modèle.

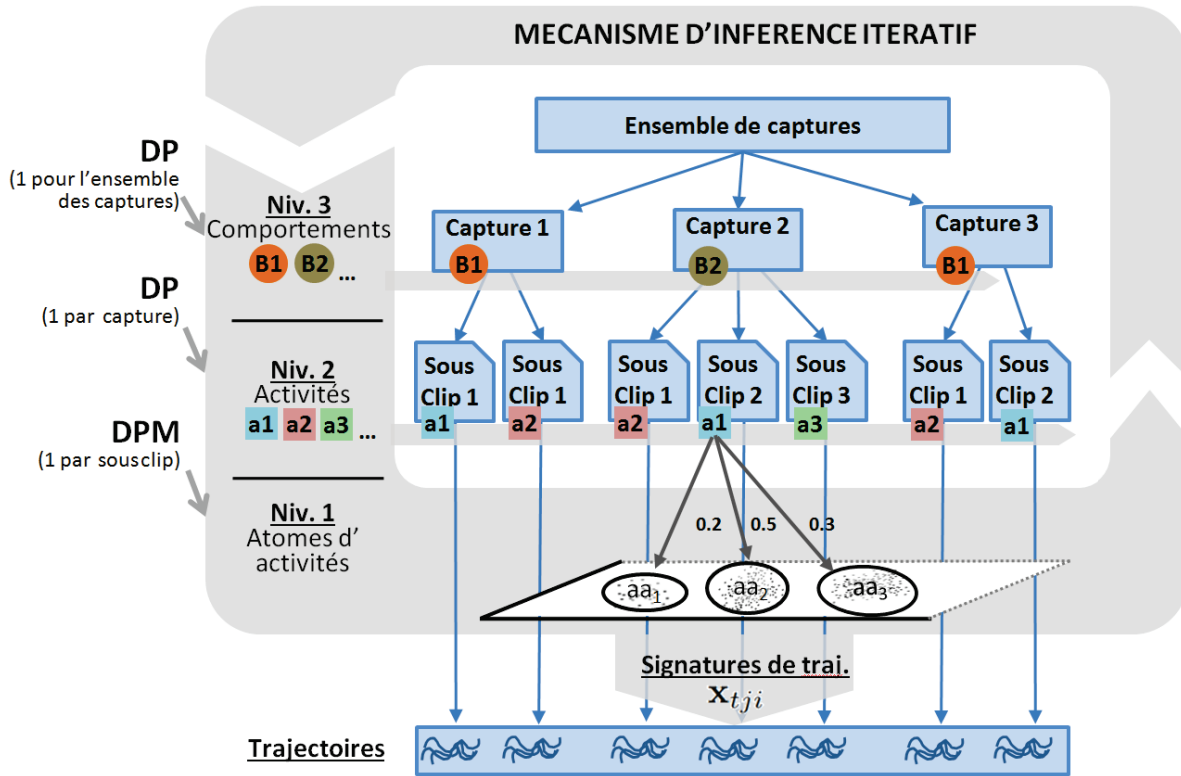


Figure 3.12 – Illustration du principe d'inférence des paramètres du modèle à partir des données. Les boîtes bleues indiquent la structure du jeu de données. Le cycle grisé représente le principe de modélisation des données selon plusieurs niveaux. Les paramètres du modèle MLC-HDP (β, π, z) s'adaptent au fur et à mesure des itérations pour expliquer au mieux les données.

3.3.3 Signatures de trajectoires

Cette section propose deux signatures permettant de représenter des trajectoires définies en 2D et 3D. Le but de nos travaux n'étant pas la recherche de signatures optimales pour la représentation de trajectoires, mais bien de trouver une signature fonctionnelle pour la mise en œuvre du MLC-HDP. Le lecteur souhaitant approfondir le domaine peut consulter [Yang 2011]. Par exemple, dans [Beyan 2013], les auteurs représentent chaque trajectoire de poisson par 179 caractéristiques (originellement 776 caractéristiques extraites, réduites par Analyse en Composantes Principales (ACP)). Les deux signatures (2D et 3D) que nous proposons par la suite s'inspirent en partie de ces travaux, notamment en considérant un mélange de deux types de caractéristiques :

- caractéristiques invariantes au contexte ;
- caractéristiques globales dépendantes du contexte.

En pratique, les trajectoires d'êtres vivants sont souvent chaotiques et nécessitent ainsi pour être qualifiées, un nombre important de caractéristiques. Contrairement à l'article mentionné précédemment, nous nous contentons d'un nombre plus modeste de caractéristiques, car la précision des clusters de trajectoires n'est pas primordiale dans notre approche. Rappelons que notre objectif est en premier lieu la classification de *captures* et *sous-clips* et non la détection de trajectoires anormales. En fait, il est difficilement envisageable d'étudier les trajectoires de

façon individuelle car il est compliqué de caractériser ce qu'est une trajectoire normale. Nous nous intéressons donc plutôt à des comportements potentiellement anormaux de cohortes. La détection de trajectoires anormales d'un point de vue individuel nécessiterait de repenser à une signature éventuellement plus riche.

3.3.3.1 Trajectoires 2D

La signature de trajectoire 2D présentée dans cette section est utilisée par la suite dans le cadre d'expériences sur des données simulées (cf. Section 3.4.3, [Chiron 2014c] et [Chiron 2014b]). Soit pour chaque trajectoire, un vecteur constitué des 7 caractéristiques invariantes ainsi que des caractéristiques globales précisées ci-dessous. En aval, à partir de l'ensemble des signatures extraites, on normalise (i.e. recentrage et division par l'écart-type) les vecteurs sur chacune de leurs dimensions.

Caractéristiques invariantes La Fonction de Distance Centrée (CDF) mentionnée dans [Bashir 2006] consiste à calculer la distance relative entre chaque point de la trajectoire et son centroïde, défini par $d_c(i) = \sqrt{(x(i) - \bar{x})^2 + (y(i) - \bar{y})^2}$ avec i l'index du point de la trajectoire. Les 7 caractéristiques suivantes sont extraites à partir du vecteur d_c : la moyenne, l'écart-type, le minimum, le maximum, l'asymétrie, le coefficient d'aplatissement et le nombre de croisements à la moyenne (en analogie avec le concept de "zero-crossing"). La figure 3.13 montre un exemple de calcul de CDF à partir d'une trajectoire 2D.

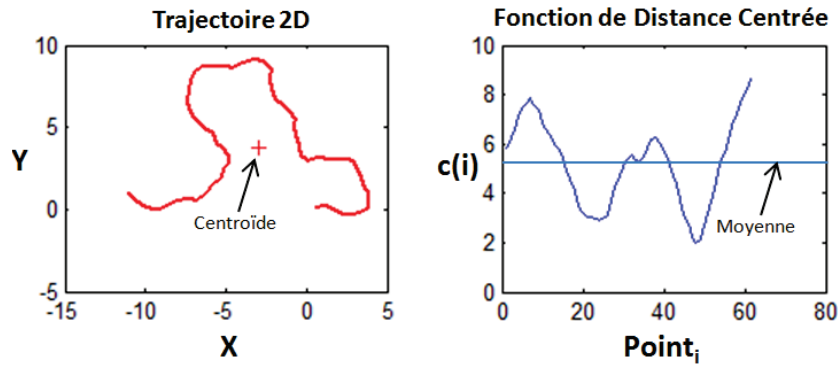


Figure 3.13 – Exemple de calcul de Fonction de Distance Centrée (CDF) à partir d'une trajectoire 2D.

Caractéristiques globales La scène 2D est définie en cellules selon une grille de n lignes sur m colonnes. Un vecteur de caractéristiques de $n \times m$ valeurs, correspondant au temps passé par la cible dans chaque cellule de la grille. Une ACP est appliquée afin de réduire le nombre de dimensions, tout en veillant à garder au moins 90% de l'information.

3.3.3.2 Trajectoires 3D

Des signatures de trajectoires 3D sont utilisées par la suite dans le cadre d'expériences sur des données simulées plus élaborées (cf. Section 3.4.4) ainsi que sur des données de trajectoires

d'abeilles réelles (cf. Chapitre 4). Pour ce type de signature, on reprend les caractéristiques invariantes qui s'adaptent facilement en 3D. En revanche, les caractéristiques dites "globales" ne peuvent être étendues dans le cas 3D du fait du nombre de cellules nécessaires pour mailler le volume de la scène. C'est pourquoi, nous préférons utiliser les caractéristiques globales suivantes : les angles moyens de la trace par rapport à l'horizontal et la vertical, ainsi que la distance moyenne à la ruche.

3.3.4 Critère d'arrêt

Le MLC-HDP explore de manière itérative le domaine des solutions avec une part d'aléatoire (approche de type Monte-Carlo, cf. construction "stick breacking", et les différentes phases d'échantillonnage). Ainsi, la méthode converge généralement vers une solution optimum locale. L'idée est alors de laisser évoluer le MLC-HDP assez longtemps et de retenir la solution (modèle de mélange de gaussiennes) qui maximise la vraisemblance par rapport aux données. Un indicateur classique est la log vraisemblance négative que l'on cherche à minimiser :

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_{n,k} N(x_n|\mu_k, \Sigma_k) \right\} \quad (3.14)$$

avec $\pi_{n,k}$ la probabilité que la donnée x_n appartienne au cluster k , et N la fonction de vraisemblance d'une loi normale multi-variée. Avec μ_k et Σ_k les paramètres de la gaussienne modélisant le cluster k . La figure 3.14 donne un aperçu de l'évolution de la log vraisemblance négative au fil des itérations du MLC-HDP. Nous avons constaté que les découvertes de nouveaux clusters pertinents se traduisent par des diminutions soudaines de la vraisemblance. Il serait possible de définir un critère d'arrêt à partir d'un seuil de stabilité de la vraisemblance sur les n dernières itérations. Nous n'avons pas encore implémenté dans le cadre de nos travaux cette seconde idée de critère d'arrêt.

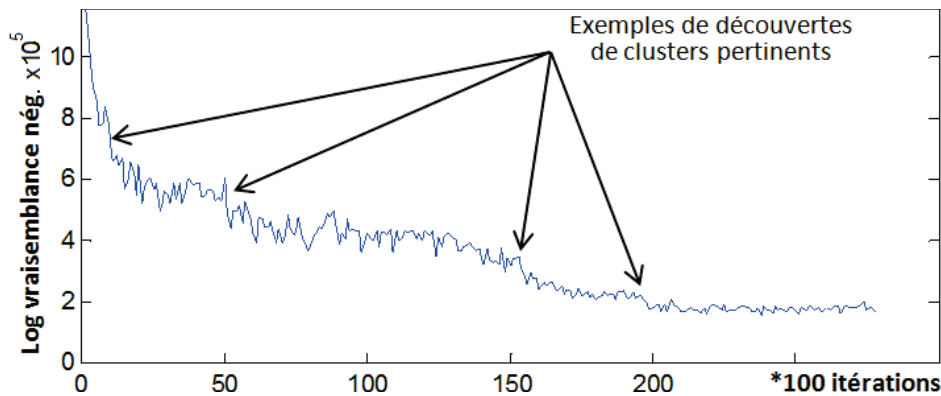


Figure 3.14 – Progression de la log vraisemblance négative au cours des itérations. Les chutes notables correspondent à la découverte de clusters pertinents.

3.4 Simulation Multi-Agents pour la validation de l'approche

3.4.1 Pertinence du couple SMA / MLC-HDP

L'avantage de la validation par un SMA est double :

Le premier avantage Les SMAs permettent de simuler des *comportements* variés à partir de modèles simples comportant peu de paramètres. On peut ainsi générer des vérités terrain sans avoir recours à un processus d'annotation laborieux. Nous le verrons par la suite, les *comportements* (concepts de haut niveau) peuvent par exemple être observés au travers de variables d'environnement (sachant que les agents agissent sur leur environnement). Quant aux trajectoires (concepts de bas niveau), elles sont directement engendrées par le déplacement des agents mobiles.

Le second avantage Les trajectoires simulées ne sont pas directement contraintes par un modèle défini de dynamique. En effet, il ne serait pas judicieux de chercher à modéliser des données dont on connaît et maîtrise en détails les paramètres permettant de les simuler. Contrairement aux scènes encombrées fréquemment rencontrées en vidéosurveillance (e.g. trafic, piétons), l'évolution d'insectes ou d'animaux en général donnent lieu à des scènes bien moins structurées. Cela s'explique par le peu de contraintes engendrées par un environnement ouvert (e.g. scène non définie par des murs, couloirs, routes). Il arrive souvent même que les cibles vivantes ne suivent aucune tendance perceptible de mouvement. Ce phénomène est aisément simulable à l'aide d'un SMA.

Finalement, l'aspect intéressant de cette approche est que les trois étapes de traitement (suivi de cibles, simulation SMA et modélisation MLC-HDP) mettent en œuvre un minimum *d'a priori* sur les données :

- le nombre de clusters pour le MLC-HDP n'a pas besoin d'être défini (cf. Section 3.3),
- la dynamique pour la prédiction (cf. Section 2.3.3) et maintenant la simulation des trajectoires peut s'appuyer essentiellement sur les données.

3.4.2 Génération du jeu de données structuré de trajectoires et de la vérité terrain associée

Lors d'une simulation SMA, chaque agent produit un nombre variables de trajectoires. On considère comme "trajectoire" la séquence de points issue du déplacement d'un agent entre un point d'intérêt A (e.g. nids, ruche) et un point d'intérêt B (e.g. source de nourriture, limite de la zone d'observation). Ces trajectoires sont collectées et rassemblées dans un jeu de données. Celui-ci est organisé selon la structure hiérarchique en *captures* / *sousclips* décrite dans la section 3.3.1. Simultanément, on peut observer certaines variables environnementales, qui serviront de vérité terrain quant à l'évolution du comportement de la colonie dans le temps. La figure 3.15 illustre comment les trajectoires sont collectées et comment des variables d'environnement peuvent témoigner du comportement de la colonie.

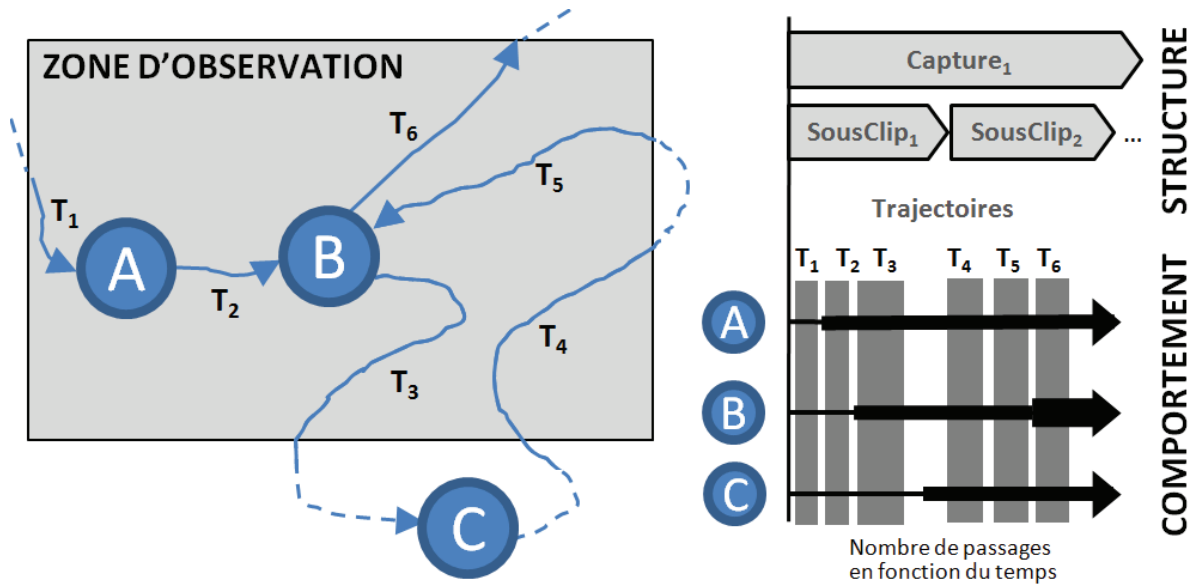


Figure 3.15 – À gauche : schéma des trajectoires engendrées par un agent à l'issue d'une simulation SMA. La collecte des données est restreinte par une zone d'observation. À droite : structure des données de trajectoires et observation du comportement au travers de variables environnementales (e.g. nombre d'arrivées aux points de référence en fonction du temps).

Niv. 3 - Captures & comportements : une *capture* correspond à une simulation complète. Celle-ci est lancée sur un jeu de paramètres et engendre un *comportement* particulier. Nous fixons arbitrairement la "durée" d'une simulation à environ 1500 itérations, soit le temps nécessaire pour que la colonie d'agents enchaîne un certain nombre d'*activités* différentes. Sinon, on peut aussi stopper la simulation lorsque le système arrive à un équilibre (e.g. ressources de l'environnement totalement consommées).

Niv. 2 - Sousclips & activités : les *captures* sont arbitrairement séparées en 32 *sousclips*. Ce qui donne des *sousclips* de $3000/32 = 92$ itérations, d'une durée suffisante pour englober un nombre significatif de trajectoires, mais pas trop pour éviter d'avoir plusieurs *activités* différentes par *sousclip*.

Niv. 1 - Signatures de trajectoires : à partir des trajectoires individuelles (2D ou 3D) sont calculées des signatures (cf. Section 3.3.3) qui serviront de données en entrée du MLC-HDP.

Dans notre cas, un modèle SMA paramétré par un jeu de paramètres fera émerger, à l'issue de différentes instances de simulation des comportements proches, voir similaires. De légères variations sont tout de même observées quant aux instants et aux vitesses de convergence des différentes activités (e.g. récolte d'une source, recherche de nourriture).

3.4.3 Étude préliminaire avec le modèle des fourmis

Dans un premier temps, en guise de preuve de concept, nous avons testé notre approche sur un modèle SMA de fourragement de fourmis [Wilensky 1997], qui est un grand classique dans la communauté multi-agents. La section suivante présente le modèle dans les grandes lignes. Puis, les deux sections qui suivent montrent les résultats de clustering à différents niveaux via la méthode MLC-HDP. Cette expérience a fait l'objet de la publication [Chiron 2014c].

3.4.3.1 Modèle de fourragement de fourmis

Les fourmis (ou agents) explorent leur environnement à la recherche de nourriture. Lorsqu'une fourmi trouve une source de nourriture, elle en rapporte un peu au nid. Le modèle prévoit que les fourmis savent retrouver leur nid. Sur le chemin du retour, la fourmi dépose des phéromones sur le sol qui vont se diffuser dans l'espace et s'évaporer avec le temps. De manière globale, les fourmis sentent les phéromones au sol et ont le réflexe de les suivre. Ceci entraîne un phénomène de renforcement des pistes menant à la nourriture. De fait, le système multi-agents converge vers un comportement (dépendant des paramètres du modèle), ce qui représente l'intérêt de cette approche pour notre analyse comportementale. Au delà des paramètres de population et d'environnement (e.g. sources de nourritures et nid), le modèle de fourragement de fourmis s'appuie principalement sur les deux paramètres suivant :

- TD : le taux (ou vitesse) de diffusion des phéromones
- TE : le taux (ou vitesse) d'évaporation des phéromones

Les variations de ces deux paramètres, TD et TE, entraînent des changements de comportement au sein de la colonie de fourmis, notamment au niveau de l'ordre, de la stabilité et de la vitesse des récoltes des sources de nourriture. La figure 3.16 illustre des simulations issues de 4 jeux de paramètres différents. L'effet des paramètres TD et TE sur le comportement de la colonie est clairement visible sur les graphiques d'évolution des réserves de nourriture. Les paramètres de simulation associées aux graphiques d'évolution de la nourriture constitueront ainsi notre vérité terrain comportementale.

Comme on peut le voir sur la figure, les configurations A (TD faible et TE moyen) et B (TD élevé et TE moyen) conduisent les fourmis à visiter les sources de nourriture une par une. À l'opposé, dans la configuration C (TD élevé, TE faible) et D (TD élevé, TE élevé) les sources sont récoltées simultanément, mais avec des degrés d'efficacité divers. En complément, la vue Netlogo apporte d'autres informations sur le comportement. Par exemple, on aperçoit, selon la configuration, un nombre plus ou moins important de fourmis qui explorent l'espace sans but précis.

On note que sur ce modèle de fourragement de fourmis, la zone d'observation s'étend sur tout l'espace de simulation, ce qui diffère quelques peu de l'application fil rouge de cette thèse où les trajectoires d'abeilles sont observées localement à l'entrée de la ruche.

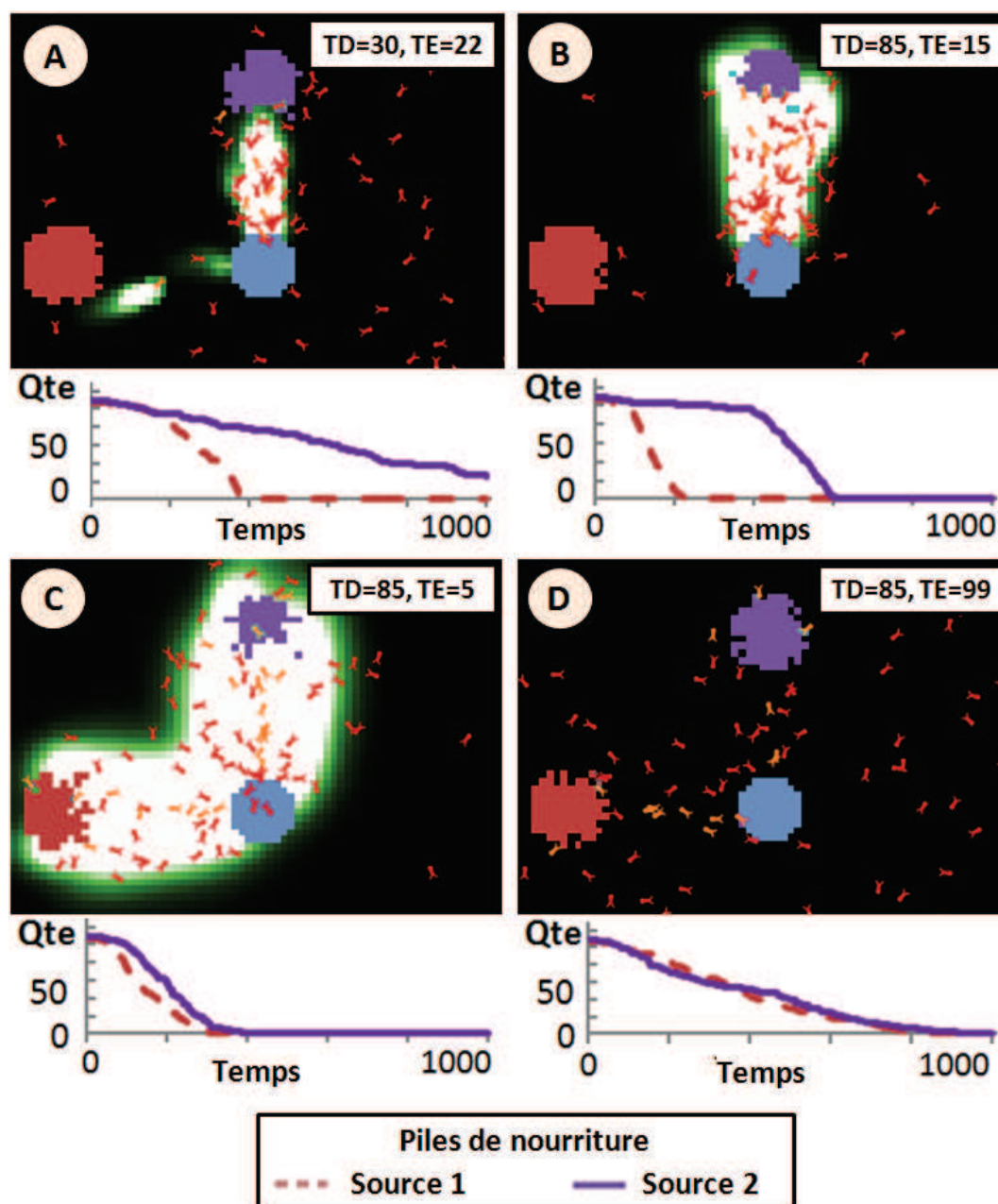


Figure 3.16 – Plusieurs simulations de l’algorithme des fourmis selon 4 configurations (A,B,C et D), chacune illustrée par une vue Netlogo (à la 150ème itération) dans la partie supérieure, et un graphique des réserves (ou sources) de nourriture dans la partie inférieure. Pour chaque vue, le cercle central représente le nid, les 2 cercles (à gauche et en haut) représentent les sources de nourriture, les points rouges symbolisent les fourmis et le gradient vert/blanc représente l’intensité des phéromones déposées au sol par les fourmis. Voici le détail des configurations : A (TD faible et TE moyen), B (TD élevé et TE moyen), C (TD élevé, TE faible) et D (TD élevé, TE élevé).

3.4.3.2 Exemples de signatures sur les trajectoires simulées

Cette section montre au travers de la figure 3.17 deux ensembles de signatures correspondant à des captures simulées selon des configurations différentes. On observe certains agrégats communs entre les deux ensembles, et d'autres complètement uniques. Cette situation se prête bien à une modélisation de type HDP, où les clusters sont partagés entre les groupes, mais représentés par des poids différents.

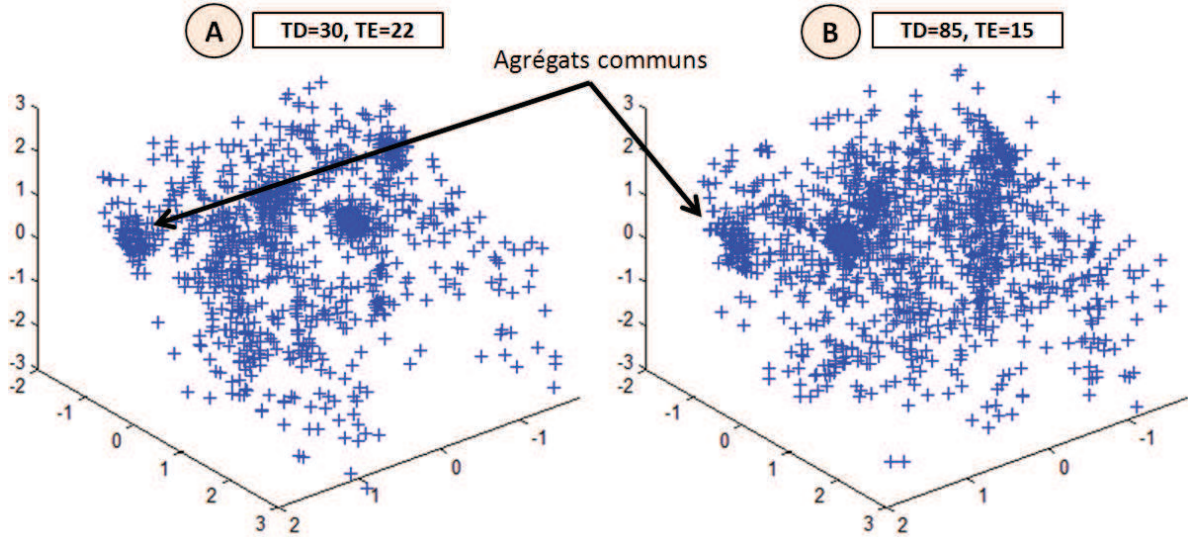


Figure 3.17 – Signatures (projetées dans un repère 3D via une ACP) pour deux capture simulées selon des configurations différentes. Le modèle multi-agents utilisé est celui des fourmis présenté dans la section précédente.

3.4.3.3 Résultats de clustering MLC-HDP - Niveau 2

Cette section illustre l'utilisation du MLC-HDP sur une seule *capture*, limitée donc au niveau 2 "*activité*". La *capture* est obtenue à partir du modèle multi-agents de fourmis défini précédemment. Les paramètres sont les suivants : TD=75, TE=10, trois sources de nourriture disposées à des distances croissantes de la ruche. La *capture* dure 1500 itérations (ou unités de temps), est segmentée en 31 *sousclips*, et comporte au total 1550 trajectoires. La figure 3.18 montre les données de simulation de la *capture* (i.e. graphique d'évolution des réserves de nourriture), et les résultats de modélisation MLC-HDP au niveau 2. L'interprétation visuelle des étapes de la chronologie démontre la pertinence des résultats de clustering. Chaque *sousclip* a été associé à une des 10 *activités* distinctes identifiées par le MLC-HDP. Ces 10 *activités* s'appuient sur 12 *atomes d'activité* auxquels sont associés différents poids en fonction de celles-ci. L'*atome d'activité* le moins, et le plus représenté, a été associé respectivement à 49, et 243 trajectoires. Comme le montre la figure 3.18, la plupart des transitions entre les *activités* sont en phase avec des événements notables s'étant produits lors de la simulation. Par exemple, sur la figure, la transition entre "Récolte source 3" et "Exploration libre" est nette, on voit bien une rupture des étiquettes (rouges et violettes) associées aux *sousclips* concernés.

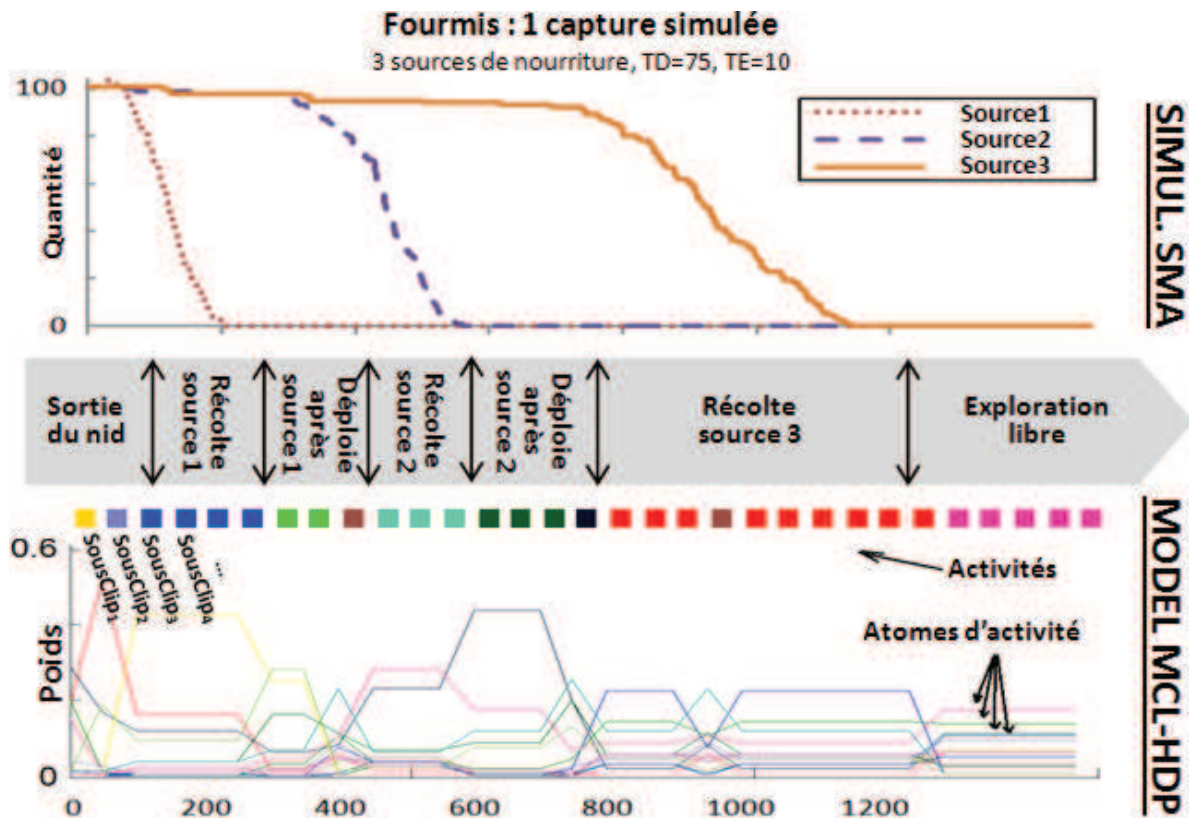


Figure 3.18 – En haut, les graphiques d'évolution des réserves de nourriture montrent que les sources de nourriture ont été récoltées une à une. En bas, les résultats de clustering MLC-HDP (niv. 2), où chaque *sousclip* est associé à une des 10 "activités" trouvées. Chaque carré symbolise, par sa couleur, l'activité correspondant au *sousclip* associé. Au centre, 7 interprétations visuelles à la série d'activités identifiées par le MLC-HDP.

Il est important de remarquer que la granularité du jeu de données (e.g. longueur des *sousclips*) peut influencer la qualité des résultats. De fait, des événements notables peuvent se mélanger au sein d'un même *sousclip* et donner naissance à des *activités* de nature différente.

3.4.3.4 Résultats du clustering MLC-HDP - Niveau 3

Cette section illustre l'application du MLC-HDP effectuée sur plusieurs *captures* permettant ainsi d'étendre l'étude jusqu'au niveau 3 "*comportement*". Le succès du clustering au plus haut niveau sémantique témoigne en partie du bon fonctionnement des couches inférieures. Dans cette expérience, les 4 *comportements* présentés précédemment (cf. Figure 3.16) sont mis en œuvre. De manière générale, le modèle des fourmis se comporte de façon identique (i.e. émergence de *comportements* similaires) lorsque que l'on réitère la simulation sur un même jeu de paramètres. Cependant, dans de rares occasions, il arrive que la simulation fasse émerger un *comportement* inattendu. Dans le cadre de cette première expérience, ces exceptions ont été manuellement retirées du jeu de données pour que la vérité terrain reste cohérente avec les *comportements* observés (i.e. paramètres de simulation / graphique d'évolution des réserves de nourriture). Finalement, 5 simulations sont retenues par jeu de paramètres, ce qui correspond à un jeu de données constitué de $4 * 5 = 20$ *captures*. L'environnement (2 sources de nourri-

tures situées à égale distance de la ruche) et la population (100 fourmis) est identique pour l'ensemble des *captures* simulées. Le jeu de données est alors composé de 49 882 trajectoires séparées en 20 *captures*, chacune segmentée en 16 *sousclips*.

La figure 3.19 montre les résultats du clustering pour les niveaux *activité* et *comportement*. La double force du MLC-HDP est clairement visible. Premièrement, 4 catégories principales ont été trouvées avec succès, le tout avec un minimum d'*a priori* sur les données, si ce n'est la structure en *capture/sousclips*. Deuxièmement, 18 sur les 20 *captures* ont été correctement classées en adéquation avec les paramètres de simulation. La couche *activité* montre la séquence d'*activités* correspondant à l'évolution temporelle de la colonie. Il est intéressant de noter que l'aspect temporel n'est pas pris en compte avec le MLC-HDP. Par exemple, le *comportement* "Récolte des sources une à une" est similaire, quelque soit l'ordre de récolte des sources, à condition que la distance des sources par rapport au nid soit identique.

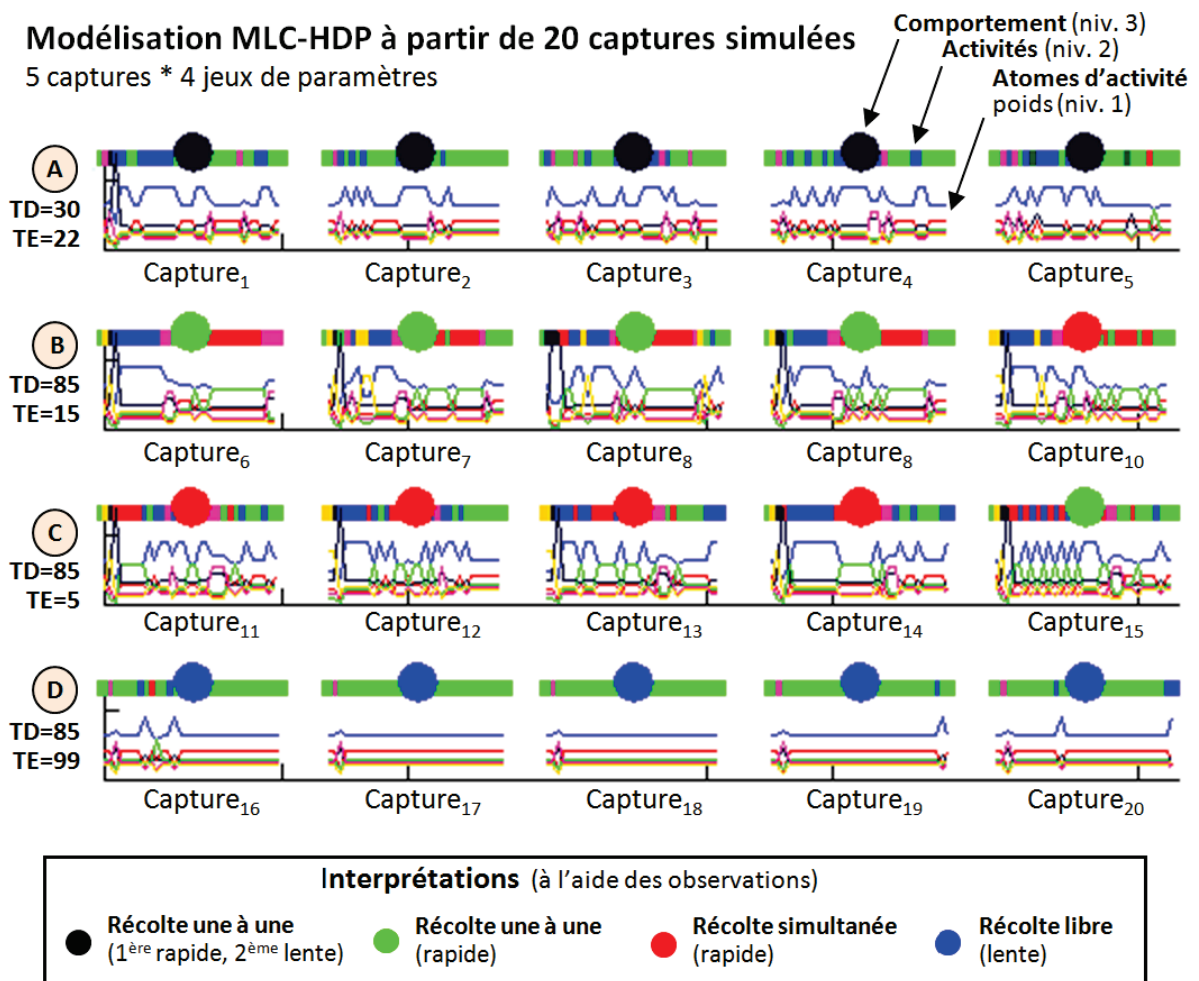


Figure 3.19 – Résultats sur 20 *captures*, chacune composée de plus de 1000 trajectoires. 4 configurations (ou jeux de paramètres) sont testées, avec 5 simulations pour chaque configuration afin d'assurer une certaine robustesse. La couleur des grands cercles au dessus de chaque *capture* représente le *comportement* associé à la *capture* (niv. 1). La couleur des petits rectangles au dessus de chaque *capture* représente l'*activité* associée à chaque *sousclip* (niv. 2). Les courbes colorées représentent pour chaque *sousclips*, les poids associés à tous les *atomes d'activité* existants (niv. 3).

Dans la section suivante est menée une étude à plus grande échelle, sur un modèle multi-agents d'abeilles évoluant en 3D. Ce dernier se rapproche plus des contraintes de notre application fil rouge.

3.4.4 Étude approfondie à partir d'un modèle de colonie des abeilles

Afin d'approfondir notre étude, nous avons implémenté en Netlogo un modèle de système multi-agents modélisant une colonie d'abeilles à la recherche de nourriture. Pour construire ce dernier, nous nous sommes inspirés de travaux tels que [Lemmens 2008] où un certain nombre de règles biologiques sont données, mais aussi [Figuerola 2009] qui propose quelques pistes pour la mise en œuvre d'un tel SMA. On retient notamment de ces travaux, la faculté des abeilles à partager les informations avec leurs congénères. Dans la nature, par exemple, l'échange d'information se fait grâce à une danse effectuée au sein de la ruche. De manière identique au modèle de fourragement de fourmis (cf. Section 3.4.3), qui converge grâce aux phéromones, notre modèle de colonie d'abeilles fait émerger des comportements grâce à ce partage d'information entre individus. Et c'est cela qui donne tout son intérêt au modèle en termes de simulation de comportements. Nous avons proposé dans [Chiron 2014b] une étude portant sur une implémentation en 2D d'un premier modèle d'abeilles. Les méthodes et résultats obtenus sont très proches de ceux présentés pour les fourmis (cf. Section 3.4.3 ou [Chiron 2014c]).

Nous proposons dans cette section d'étudier plus en détails les résultats du MLC-HDP vis-à-vis de données qui, nous l'espérons, reflètent mieux la réalité. Pour cela, notre étude utilise sur une nouvelle implémentation du modèle de colonie d'abeilles dans un contexte 3D, produisant des trajectoires plus bruitées et générant des comportements plus complexes. Cela dans le but de se rapprocher de la réalité, et valider ainsi notre démarche méthodologique en vue d'établir des protocoles expérimentaux robustes (cf. Chapitre 4).

3.4.4.1 Modèle de colonie abeilles proposé

Voici en quelques lignes une présentation du fonctionnement de notre modèle de colonie d'abeilles. Le lecteur peut se référer à l'annexe B pour plus de détails. Chaque abeille de la colonie est représentée par un agent. Une abeille peut prendre tour à tour les 3 statuts suivants :

- **Exploratrice** : explore le monde à la recherche d'une source de nourriture. Si l'exploration demeure infructueuse au delà d'un laps de temps donné, l'abeille est contrainte de revenir à la ruche. Lorsque qu'une source de nourriture est trouvée, l'abeille peut devenir "Exploitante" ;
- **Exploitante** : exploite une source de nourriture en faisant des allers-retours entre la source et la ruche. Lorsque la source est épuisée, l'abeille peut redevenir "Exploratrice" ou devenir "Suiveuse" ;
- **Suiveuse** : positionnée dans la ruche, elle accumule les informations sur les sources de nourriture partagées par les "Exploitantes" lors des multiples va-et-vient de celles-ci. Au bout d'une certaine période, la "Suiveuse" peut devenir "Exploitante" et exploiter à son tour la source de nourriture la plus partagée par ses congénères.

La figure 3.20 représente l'automate et les conditions de passage d'un état à l'autre.

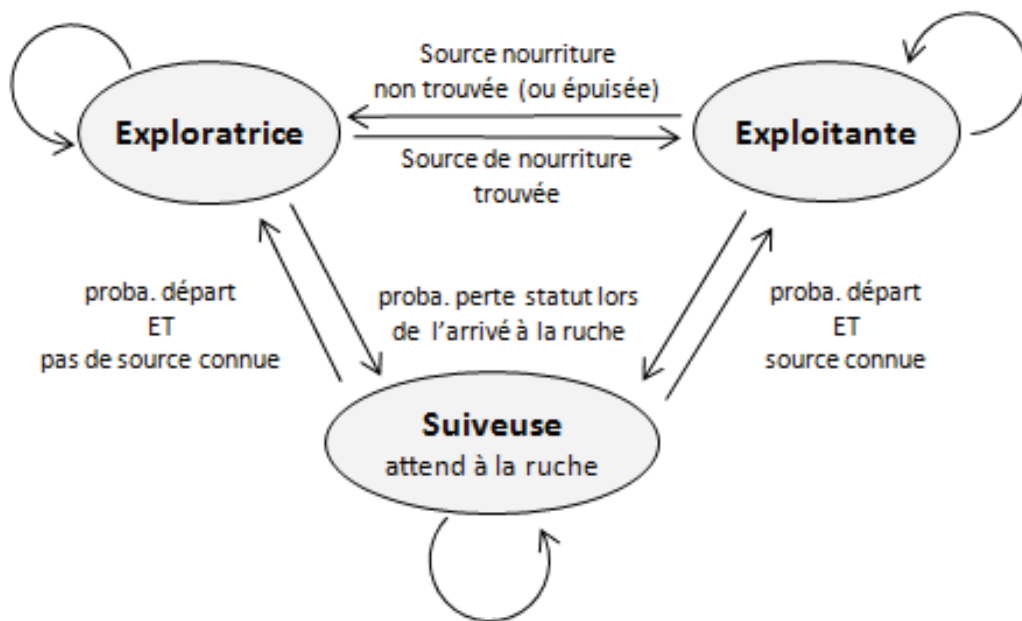


Figure 3.20 – Automate modélisant les conditions de transition entre les 3 statuts que peuvent prendre les agents. L'algorithme (cf. Annexe B) détaille le modèle et les paramètres de manière plus précise.

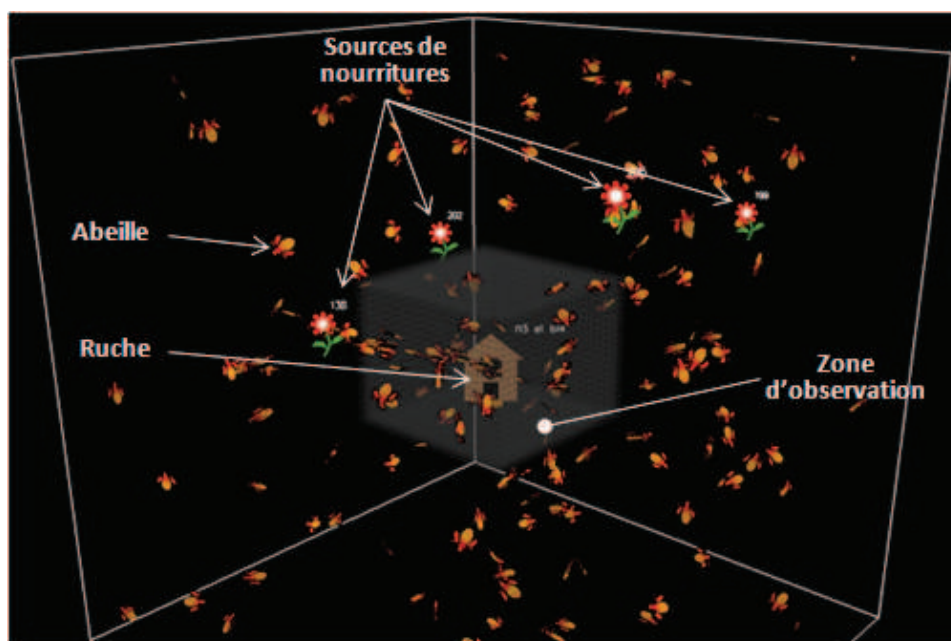


Figure 3.21 – Vue 3D Netlogo d'une simulation en cours (300ème itération) de notre modèle multi-agents de colonie d'abeilles. Le symbole "maison" au centre représente la ruche, les 4 fleurs représentent les 4 sources de nourriture, les patches oranges/rouges symbolisent les abeilles, et la zone grisée/transparente délimite la zone d'observation.

Nous proposons de nous focaliser sur l'émergence de comportements liés à des conditions d'intoxication, de prédation. Dans ce but, nous axons notre modèle autour de deux paramètres principaux que nous ferons varier, et qui sous-entendent des changements comportementaux :

- TC : le taux de communication entre les abeilles au niveau de la ruche ;
- TS : le taux de sensibilité (ou de détection) des abeilles vis-à-vis des sources de nourriture rencontrées.

D'autres paramètres de second plan sont fixés par défaut, par exemple : le nombre d'abeilles (ou population) ; la position de la ruche ; le nombre, la position des sources de nourriture et la quantité associée ; la probabilité de perdre son statut lors de l'arrivée à la ruche ; la fréquence maximum des départs au niveau de la ruche. Dans notre modèle, plus la source de nourriture est éloignée de la ruche, plus l'abeille sera rapide et sa trajectoire bruitée. Ceci modélise notamment de l'incertitude de la direction prise au départ de la ruche et permet, par la même occasion, de générer des trajectoires aux caractéristiques intrinsèques variées, contrairement aux modèles SMA précédents. L'algorithme mis en œuvre dans ce modèle est détaillé dans l'annexe B. La figure 3.21 illustre ce modèle en action. Contrairement aux modèles précédents, nous avons restreint la zone d'observation à un périmètre limité autour de la ruche, ce qui se correspond mieux aux conditions réelles, pour lesquelles, de fait, les trajectoires sont observées uniquement lorsqu'elles passent dans le champ de la caméra, elle-même placée à proximité de la ruche.

3.4.4.2 Résultats de clustering MLC-HDP - Niveau 2

De manière identique aux études précédentes utilisant une première version du SMA (i.e. fourragement de fourmis), nous nous intéressons en premier lieu aux résultats de la modélisation MLC-HDP sur une seule *capture*. Dans l'exemple qui suit, les paramètres utilisés pour la simulation SMA sont : population=200, TS=80, TC=80, quatre sources de nourriture disposées à une distance croissante de la ruche. La *capture* dure 2365 itérations (ou unités de temps), est segmentée en 30 *sousclips*, et comporte en tout 4058 trajectoires. La figure 3.22 illustre l'ensemble des trajectoires 3D contenues dans chaque *sousclip* de la *capture*.

Concernant les résultats, chaque *sousclip* a été associé à une des 11 *activités* distinctes recensées par le MLC-HDP. Ces 11 *activités* s'appuient sur 16 *atomes d'activité* auxquels sont associés différents poids. Suivant la même représentation que dans les modèles précédents, la figure 3.23 montre les étiquettes d'*activités* obtenues pour chaque *sousclip*, ainsi que l'interprétation que l'on peut en faire en nous appuyant sur la vérité terrain.

La figure 3.24 illustre les résultats de clustering des signatures de trajectoires en "*atomes d'activité*" (niv. 1). Sur cet exemple, on remarque la capacité du MLC-HDP (grâce aux DPs et DPMs) à expliquer les données par un mélange de lois normales multivariées (représentées sur la figure par des gaussiennes tridimensionnelles).

Aux quatre gaussiennes les plus importantes sont associées des trajectoires dites de "bruit" (i.e. ne correspondant à aucune activité clairement distinguable), ce qui correspond probablement à des trajectoires d'abeilles en mode "exploration". En revanche les supports des autres gaussiennes correspondent aux zones où la concentration en signatures de trajectoires est forte, et donc l'activité associée est moins chaotique. On note d'ailleurs que les paramètres des gaussiennes (μ et σ) couvrent un domaine large.

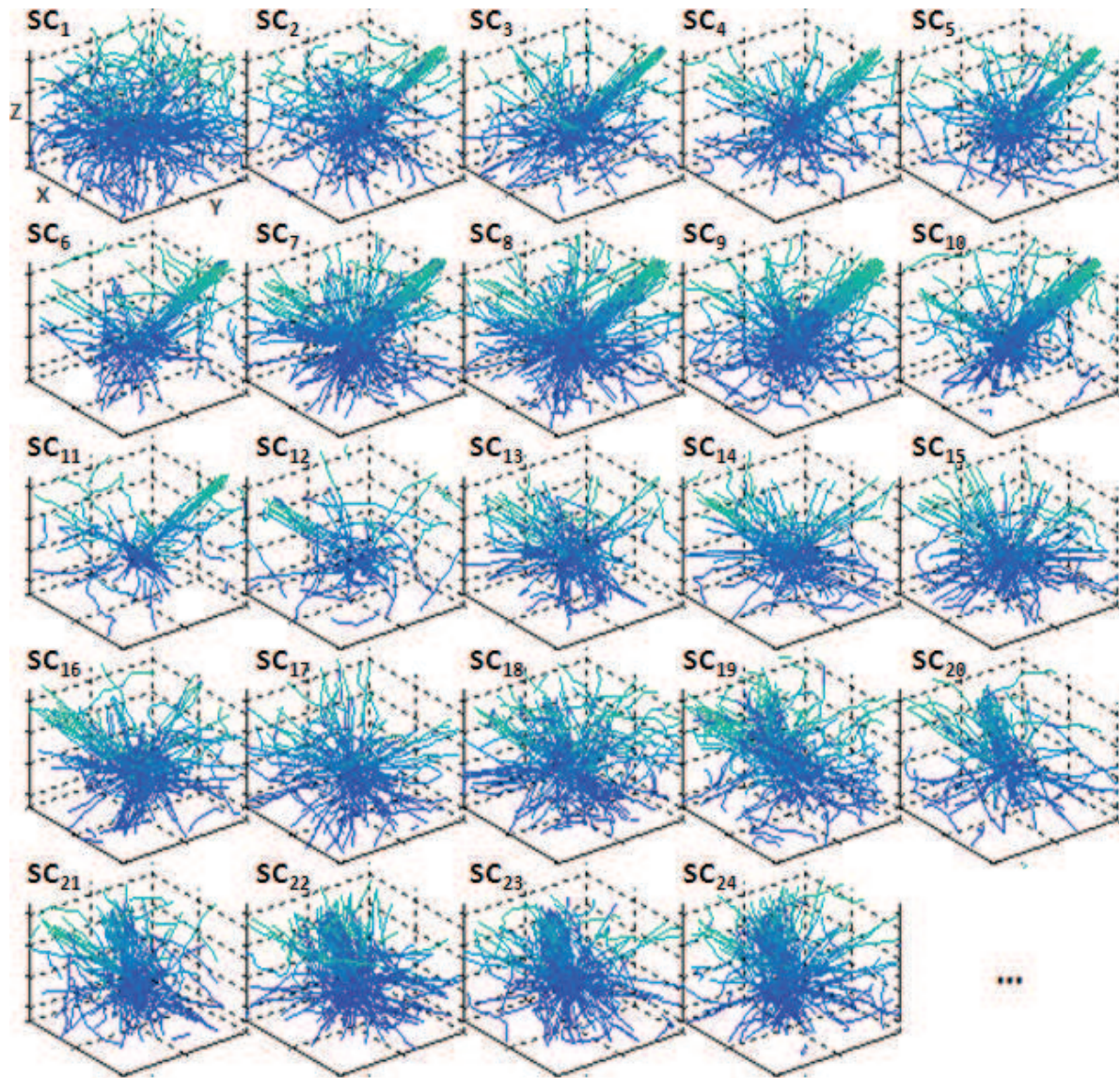


Figure 3.22 – Ensemble des trajectoires 3D contenues dans chaque *sousclips* (SC_n) d'une *capture*. Notons la ressemblance avec les trajectoires récoltées dans le chapitre 2. Le gradient bleuté indique la profondeur de la trace sur l'axe Z.

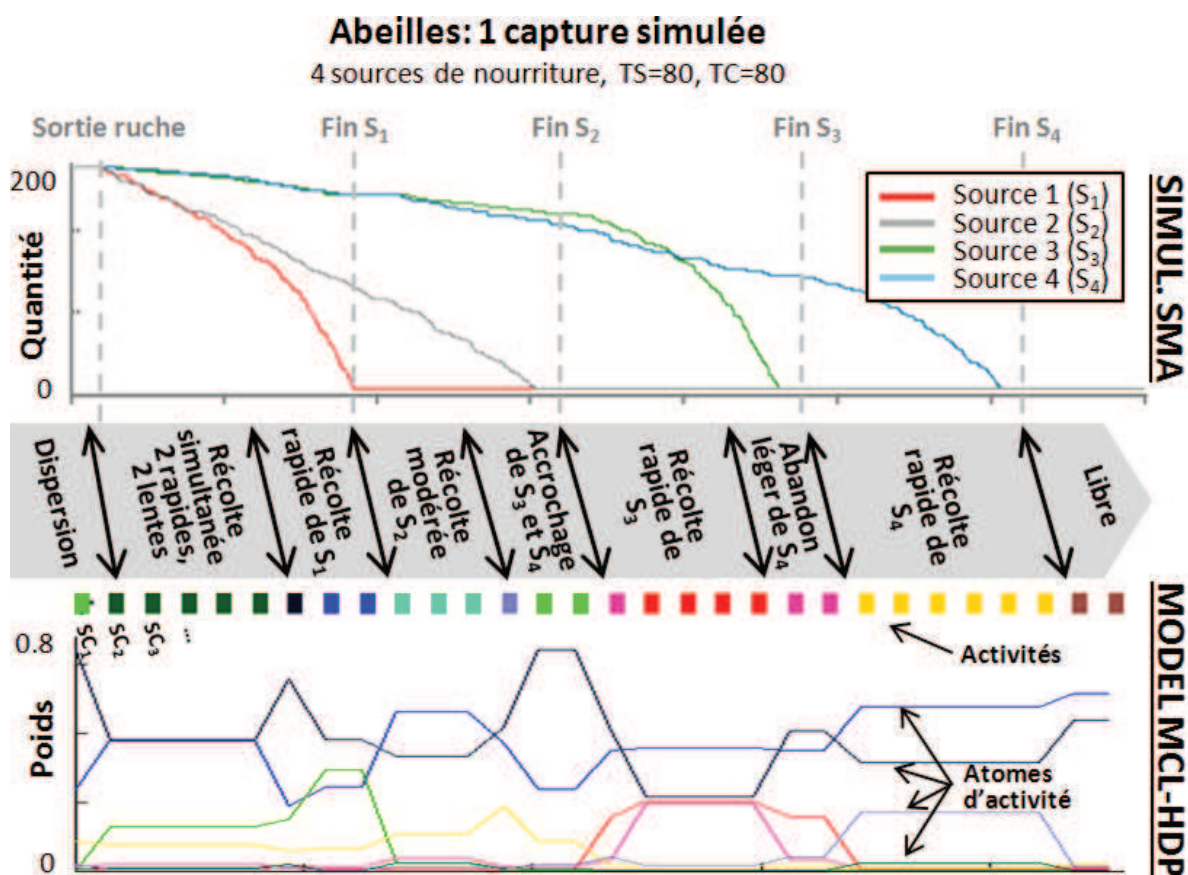


Figure 3.23 – En haut, graphiques d'évolution des sources de nourriture. En bas, résultats du clustering MLC-HDP (niv. 2). Au centre, 9 interprétations à la série d'activités obtenues par le MLC-HDP. Le code couleur des "atomes d'activité" est commun à la figure 3.24, où ils sont représentés par des clusters sous forme d'ellipsoïdes.

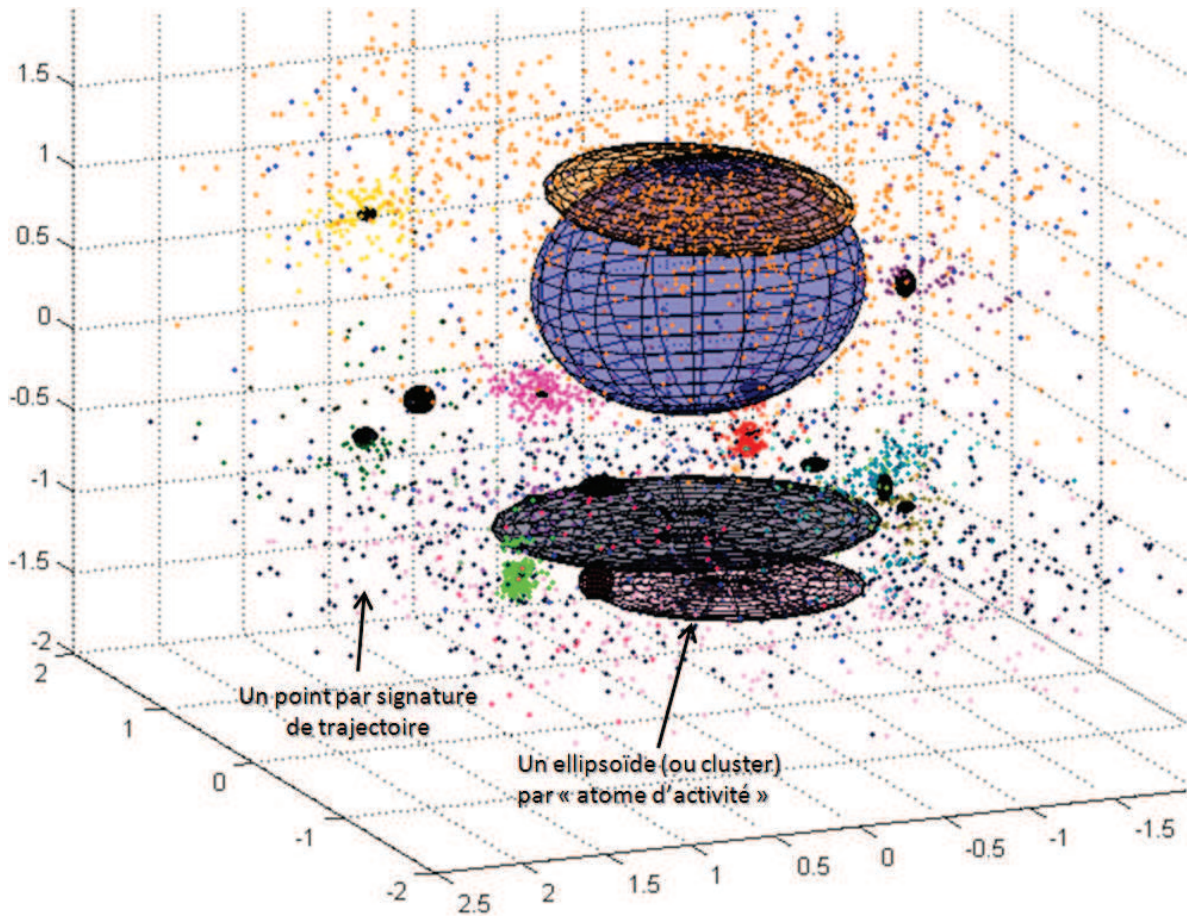


Figure 3.24 – Résultat de clustering (niv. 1) sur l'ensemble des signatures de trajectoires contenues dans une *capture*. Les 16 ellipsoïdes représentent les clusters, et la couleur indique le cluster d'appartenance des points. Une ACP a été appliquée sur les données de manière à réduire le nombre de dimensions à trois afin de permettre une visualisation pertinente.

3.4.4.3 Résultats du clustering MLC-HDP - Niveau 3

Nous testons dans cette section la capacité du MLC-HDP à regrouper un grand nombre de *captures* en fonction de leur *comportement* (niv. 3). Pour cela, nous avons simulé 80 *captures* à partir de 4 configurations de paramètres différents, soit 20 *captures* par configuration. En comparaison avec les modèles de SMA précédents (cf. Section 3.4.3 et [Chiron 2014b, Chiron 2014c]), ce modèle de colonie d'abeilles donne lieu à un plus grand nombre de *captures* "déviante", autrement dit, la variabilité du jeu de données pour une configuration (jeu de paramètres) donnée est relativement élevée. Il est à noter que dans la cadre de cette expérience, l'ensemble des *captures* (plus ou moins représentatives des paramètres du modèle) ont été conservées. Voici les 4 jeux de paramètres que nous avons fait varier dans notre modèle :

- **Configuration A (TC=95, TS=35)** : récolte lente car peu sensible aux sources de nourriture, mais convergence (i.e. effort de la colonie focalisée sur une ou plusieurs tâches utiles) facilitée par un taux de communication élevé.
- **Configuration B (TC=1, TS=35)** : récolte lente car peu sensible aux sources de

nourriture, et aucune convergence vers l'une ou l'autre des sources.

- **Configuration C (TC=1, TS=95)** : récolte rapide car très sensible aux sources de nourriture, mais aucune convergence vers l'une ou l'autre des sources.
- **Configuration D (TC=95, TS=95)** : récolte rapide car très sensible aux sources de nourriture, et convergence facilitée par un taux de communication élevé.

Les paramètres communs aux 4 configurations sont, une population de 200 individus et 2 sources de nourriture.

La figure 3.25 illustre à la fois le jeu de données (vérités terrains associées aux 80 *captures*) et les résultats de clustering du MLC-HDP sur les *comportements* (niv. 3). En termes de résultat, le MLC-HDP a découvert 8 *comportements* différents (clusters de niv. 3), 16 *activités* (clusters de niv. 2) et 6 *atomes d'activité* (clusters de trajectoires, niv. 1). Parmi les 8 *comportements*, on perçoit clairement 4 *comportements* majeurs qui correspondent aux 4 jeux de paramètres utilisés pour la simulation des *captures*. Les 4 autres *comportements* représentés de façon minoritaire permettent probablement de modéliser les comportements déviants issus de la grande variabilité du jeu de données.

Nous utilisons dans cette expérience un jeu de données constitué de 224 309 signatures de trajectoires. Celui-ci est structuré en *sousclips* et *captures* auxquels sont associés des Processus de Dirichlet. Ce qui revient à échantillonner 2480 DP par itération. Nous avons lancé le MLC-HDP sur 810 itérations, ce qui correspond à environ 14h de calcul sur une machine équipée d'un Intel Core i3 à 2.5 Ghz.

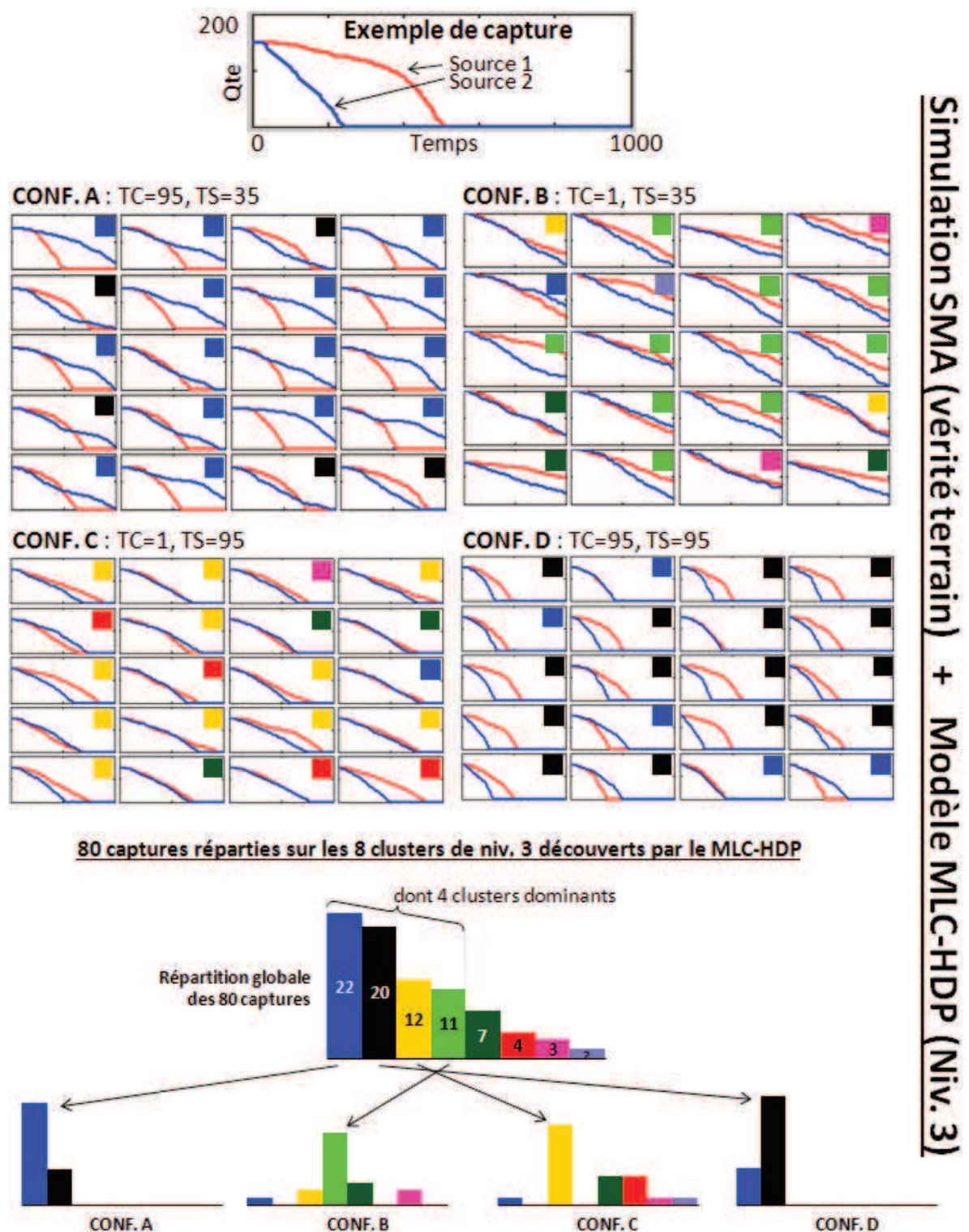


Figure 3.25 – Données de simulation multi-agents "Abeilles" et résultats du clustering MLC-HDP (niv. 3). Au premier plan, est présentée la vérité terrain comportementale (jeux de paramètres & graphiques d'évolution des réserves de nourriture) pour les 80 *captures* générées à partir de notre modèle multi-agents de colonie d'abeilles. Au second plan est présenté les résultats de la modélisation MLC-HDP au niveau 3, "*comportement*", à partir des 224 309 trajectoires structurées en *captures*.

3.5 Conclusion

Dans ce chapitre, nous avons proposé une approche pour la découverte et la classification de comportements à partir d'un ensemble structuré de trajectoires selon différentes échelles temporelles. L'intérêt principal de l'approche est qu'elle ne nécessite pas d'injection de connaissances métiers. Et finalement c'est la structure de données elle même qui fournit l'*a priori* nécessaire pour l'élaboration du modèle.

Ensuite, dans ce contexte de minimisation de l'*a priori*, nous avons proposé une validation à l'aide de modèles multi-agents. Les expériences menées à partir de données simulées ont confirmé l'intérêt d'une étude comportementale multi-niveaux sur des données structurées hiérarchiquement .

Finalement, les contributions apportées par ce chapitre sont les suivantes :

- l'utilisation d'un HDP multi-niveaux pour la découverte de comportements émergents. À notre connaissance, les approches de type HDP, en relation avec la modélisation comportementale, utilisent au maximum 2 niveaux ;
- l'utilisation de caractéristiques extraites de trajectoires individuelles éparses, plutôt qu'une approche fondée sur le flot optique telle que traditionnellement adoptée pour les scènes encombrées ;
- la validation originale par simulation SMA, fournissant indirectement la vérité terrain sémantique associée au jeu de données de trajectoires.

En termes de perspective, il serait intéressant de prendre en compte l'évolution des trajectoires (ou plutôt de leurs caractéristiques) dans le temps en étudiant les probabilités de transition aux niveaux des *activités* et des *comportements*. Dans ce contexte, les méthodes de type HDP-HMM (mentionnée dans la section 3.2.3.3) sont une piste à explorer.

Il ne reste maintenant plus qu'à confronter notre méthodologie sur une application réelle, pour ainsi mieux définir les protocoles expérimentaux nécessaire à une étude comportementale robuste. Ceci sera l'objet du chapitre 4.

Applications à la surveillance apicole dans le cadre d'études biologiques

Sommaire

4.1	Introduction	120
4.2	Dispositifs existants pour la surveillance de ruches	121
4.2.1	Compteurs d'entrées/sorties	121
4.2.2	Analyse des trajectoires	123
4.2.3	Autres approches	124
4.3	Protocole expérimental	124
4.3.1	Contraintes de terrain	125
4.3.2	Calendrier des interventions	126
4.3.3	Dispositif d'acquisition	127
4.3.4	Préparation des données et prétraitements	128
4.4	Traitement des séquences vidéo	130
4.4.1	Détection des abeilles	130
4.4.2	Calcul des trajectoires	132
4.4.3	Extraction des caractéristiques	132
4.4.4	Modélisation MLC-HDP pour RISQAPI	135
4.4.5	Synthèse sur les performances	139
4.5	Identification des trajectoires d'abeilles	141
4.5.1	Mise en correspondance des histoires de vie avec les trajectoires	141
4.5.2	Détection visuelle des puces RFID	142
4.5.3	Nettoyage des données RFID	143
4.5.4	Modalités d'application pour EPERAS	145
4.6	Travaux annexes sur la détection du frelon asiatique	146
4.6.1	Détection du frelon par sa taille	146
4.6.2	Analyse de l'activité sur la planche d'envol	147
4.7	Conclusion	148

4.1 Introduction

À l'issue des 3 chapitres précédents, nous avons présenté une chaîne complète d'acquisition et de traitement constituée des blocs suivants : 1) acquisition vidéo, 2) détection de cibles, 3) calcul des trajectoires individuelles et 4) modélisation comportementale. Ce quatrième chapitre se présente comme une étude de faisabilité concernant l'intégration de notre chaîne complète dans le cadre d'études biologiques réelles. En particulier, nous proposons ici d'identifier et de lever les derniers verrous nécessaires pour la validation de notre méthodologie.

Le projet EPERAS¹ (financé par la région) est un projet interdisciplinaire d'étude toxicologique visant à mesurer les effets de l'exposition — à des doses sublétales — de produits phytosanitaires sur l'abeille domestique. Dans le cadre du lancement du projet EPERAS (et d'un projet complémentaire soutenu par la FREDD), nous avons mis en place, en collaboration avec l'INRA Magneraud, une campagne de collecte de données de test à moyenne échelle. Ainsi, suivant un protocole rigoureusement établi (i.e. cohortes intoxiquées à différents niveaux), nous avons filmé pendant plus de 20 heures cumulées une colonie expérimentale d'abeilles à l'entrée d'une ruche.

Le projet RISQAPI² (labellisé ANR) vise entre autres à mesurer l'impact sur les abeilles de la prédation par le frelon asiatique. Ce projet s'inscrit dans le cadre d'un second partenariat avec le Muséum National d'Histoire Naturelle et l'INRA d'Avignon. Ce chapitre propose également des tests en rapport avec cette thématique. Cela dit, les travaux d'application de notre chaîne de traitement sur des données vidéo réelles dédiées au frelon sont encore en cours de finalisation à l'heure actuelle (e.g. distinction frelon/abeille fonctionnelle ; calcul du taux de prélèvement à l'étude).

Problématique Nous abordons dans ce chapitre les difficultés inhérentes à l'organisation et la mise en œuvre du protocole expérimental, et notamment les compromis qu'il a fallu trouver pour satisfaire les besoins et les contraintes de chacun des deux parties (informatique & biologie). Aussi, les expériences menées sur le terrain en conditions réelles nous ont permis d'identifier de nouveaux verrous techniques et scientifiques (e.g. fusion de données multi-source : vidéo/RFID). Ces verrous sont intéressants, aussi bien d'un point de vue des opportunités qu'ils offrent une fois levés, que par rapport aux problématiques de traitement de données qu'ils soulèvent.

Littérature La littérature info./bio. récente montre que les biologistes ont recours à des outils de plus en plus sophistiqués pour la collecte et l'analyse de données. À l'heure où les études se doivent de s'appuyer sur une quantité de données toujours plus importante, tout en ayant des contraintes de coût toujours plus fortes, la vision par ordinateur est une approche de plus en plus plébiscitée par les biologistes. Certains géants comme Microsoft Research proposent d'ailleurs des solutions clés en main telles que ZooTracker³ pour l'étude du vivant via la collecte de trajectoires à différentes échelles. On note aussi la récente émergence de la

1. Développement d'un système utilisant l'abeille pour la surveillance environnementale de sites à risque technologique (<http://l3i.univ-larochelle.fr/EPERAS>)

2. Lien spatial entre usage des pesticides et risques d'affaiblissement des colonies d'abeilles en système de grandes cultures (<http://l3i.univ-larochelle.fr/RISQAPI>)

3. Zootracer est un logiciel développé par Microsoft Research, capable de suivre individuellement de multiples individus non-marqués à partir de vidéos. Méthode fondée sur [Buchanan 2006].

communauté VAIB⁴, ainsi que l'apparition d'éditions spéciales de journaux internationaux (e.g. EURASIP JIVP⁵, special issues on "Animal and Insect Behaviour Understanding in Image Sequences"). Ces supports font l'objet d'un nombre grandissant de contributions.

Approche proposée En termes de contributions, ce chapitre 4 propose un protocole expérimental et détaille les méthodes nécessaires à la mise en place de notre dispositif d'analyse comportementale en conditions réelles. L'étude comportementale des abeilles à partir de leurs trajectoires de vol devant la ruche est complètement novatrice. Par ailleurs, ce chapitre montre comment l'implication forte de chercheurs informaticiens dans des projets de recherche relevant du domaine de la science du vivant a amené à réfléchir et à développer des briques supplémentaires de détection et de suivi d'objets, et à proposer une méthodologie d'analyse.

4.2 Dispositifs existants pour la surveillance de ruches

De nos jours, le suivi (au sens biologique) des ruches nécessite une intervention humaine régulière. En effet, les apiculteurs et les biologistes s'appuient sur des outils tels que des balances pour mesurer le poids des colonies, ainsi que différents procédés d'analyses chimiques issues de prélèvements sur le terrain (e.g. miel, cadavres d'abeilles). L'activité (au sens quantité d'abeilles actives) de la colonie peut également être estimée de manière visuelle, voir simplement à l'oreille. Des chercheurs et ingénieurs ont mis au point différents dispositifs dans le but d'accompagner, voir remplacer complètement, les apiculteurs dans certaines de ces tâches laborieuses. Cette section propose un bref aperçu des solutions automatisées existantes pour le suivi de colonies d'abeilles. Certaines sont fonctionnelles, d'autres sont encore au stade de la recherche.

4.2.1 Compteurs d'entrées/sorties

Pour les biologistes, le nombre d'entrées/sorties d'abeilles au niveau de la ruche est un indicateur qui s'avère pertinent en termes de santé de la colonie. Depuis le début du siècle, plusieurs approches de comptage s'appuyant sur différentes technologies ont été proposées. On distingue trois d'approches :

- avec identification individuelle des abeilles lors de leurs passages au niveau du dispositif de comptage ;
- avec identification dite "de cohorte" ;
- sans identification.

4.2.1.1 Avec identification individuelle des abeilles

Les compteurs avec identification individuelle permettent d'établir une histoire de vie pour chaque abeille. Par exemple, il est possible de trouver directement la fréquence des entrées/sorties de l'abeille, ou l'âge de son premier/dernier voyage à l'extérieur de la ruche. Ainsi, en période de butinage, il est également possible d'évaluer la distance qui sépare une

4. Visual observation and analysis of Vertebrate And Insect Behavior. Communauté se regroupant lors des Workshops biannuels organisés par Pr. Robert Fisher en marge de la conférence ICPR depuis 2008.

5. EURASIP Journal on Image and Video Processing est édité par Springer en mode "open access".

abeille de sa source de pollen, en s'appuyant sur une valeur de vitesse moyenne des abeilles en vol.

Ce genre de dispositif permettant l'identification des abeilles requiert une intervention préalable (e.g. pose d'un tag) sur chaque abeille à suivre. Sachant qu'une colonie compte plusieurs milliers d'abeilles, il est d'usage de n'équiper et d'étudier qu'un échantillon de la colonie.

En 1989, à l'initiative du Dr. Buchmann, la société Data2 conçoit un système de code-barres miniature pour répondre à la problématique de comptage et d'identification individuelle des abeilles. Le code-barres collé sur le thorax pèse 20 fois moins lourd que le pollen porté par l'abeille. Les abeilles sont canalisées dans un couloir à l'entrée de la ruche où les entrées/sorties sont enregistrées par un lecteur de code-barres laser.

L'INRA — Avignon et Magneraud — et depuis peu, une équipe de chercheurs australiens⁶, utilisent un système RFID miniaturisé pour étudier l'impact des pesticides sur les abeilles (cf. [Decourtye 2011]). De la même manière que pour les codes-barres, chaque abeille est équipée d'une puce RFID miniature associée à un code unique. À l'entrée de la ruche, les abeilles sont canalisées successivement au travers de 2 rangées de tunnels de 20*5mm chacun surplombé par un lecteur RFID. Les abeilles sont ainsi détectées 2 fois lors d'un passage, ce qui permet de calculer la direction — entrée ou sortie — de ce dernier. Selon [Streit 2003], ce système atteint une précision supérieure à 99.6% sur un échantillon de 300 vols. Nous verrons dans la section 4.5.3 que dans le cadre de notre application, nous serons confrontés à des données beaucoup plus bruitées. On note que la durée de vie des abeilles étant de quelques semaines, les puces RFID sont rapidement disséminées dans la nature.

4.2.1.2 Avec identification de cohorte

L'équipe "Écologie des invertébrés" de l'INRA d'Avignon a mis au point deux types de compteurs dits "de cohorte". Le premier type fonctionne à l'aide de pastilles métalliques collées sur le thorax des abeilles. À l'entrée de la ruche, les abeilles sont canalisées sous un détecteur de métal miniature qui collecte les entrées/sorties de celles équipées d'une pastille. Cette approche permet de suivre une seule cohorte. Le second type de compteur, plus évolué, s'appuie sur des pastilles colorées. Ce dernier permet d'étudier simultanément les entrées/sorties d'abeilles issues de plusieurs cohortes. Une caméra, utilisée en situation contrôlée, et des méthodes de seuillage finement paramétrées permettent de distinguer la couleur des pastilles. Ce dispositif a été validé sur 3 cohortes, ce qui correspond à 3 couleurs différentes [Brillet 2002].

Ces approches de comptage par cohortes permettent de mettre en œuvre des protocoles expérimentaux, tels que l'étude de l'influence de différents niveaux d'intoxication, simultanément sur un grand nombre d'abeilles.

4.2.1.3 Sans identification

Moins contraignants, puisque ne nécessitant pas d'intervention au préalable sur la colonie, les systèmes de comptage sans identification permettent de compter de manière globale les entrées/sorties des abeilles en un point donné. Dans les années 30, [Chauvin 1976] introduit un système de comptage totalement mécanique. Plutôt intrusif, ce principe mécanique cède rapidement sa place à des systèmes de comptage reposant sur des cellules photoélectriques

6. Projet "Swarm sensing" dirigé par Dr. Paulo de Souza, directeur du "Commonwealth Scientific and Industrial Research Organisation" (CSIRO).

(e.g. [Spangler 1969]). Ces derniers nécessitent un entretien lourd et souffrent par ailleurs pour la plupart de problèmes de précision. En 1994, [Struye 1994] introduit un dispositif de comptage infrarouge relativement robuste et efficace, mais la canalisation des abeilles dans des couloirs s'avère toutefois envahissante et perturbe localement le comportement des abeilles.

Le prototype présenté dans des travaux de recherche [Campbell 2008] utilise une caméra classique (RGB, 640*480 px et 30 Hz). Il se démarque des autres systèmes par son caractère non intrusif, que ce soit au niveau de la ruche ou bien au niveau des abeilles. La caméra est placée à une hauteur de 20 cm de la planche d'envol, le champ de vision horizontal de 80° permet d'observer l'entrée de la ruche (40 cm de large) dans son intégralité.

En 2011, après la réalisation d'un premier prototype par une équipe de l'INRA d'Avignon, une collaboration entre le L3i, l'INRA et la société ApiLab⁷ s'inspire de ce prototype et donna naissance à une nouvelle méthode de comptage d'abeilles, utilisant aussi la vidéo (cf. rapport interne [Blois 2011]). C'est ce système et ces algorithmes qui sont utilisés dans la plateforme actuellement commercialisée par la société ApiLab.

4.2.2 Analyse des trajectoires

Il existe dans la littérature scientifique des études biologiques ayant recours à des méthodes de calcul de trajectoires d'abeilles. La plupart permettent seulement de capturer des trajectoires d'abeilles en 2D. Voici une liste non exhaustive de références bioinformatiques.

En condition laboratoire Par exemple les auteurs de [Williamson 2013] ont démontré l'effet d'un pesticide sur le comportement des abeilles en condition laboratoire. Pour cela, ils utilisent "The Observer XT", un outil commercialisé par Noldus qui permet de suivre des insectes en conditions contrôlées. De son côté, [Kimura 2014] propose une nouvelle méthode pour le suivi de multiples abeilles en 2D dans des boîtes transparentes plates.

À l'intérieur de la ruche C'est pour suivre des abeilles à l'intérieur de la ruche que le plus grand nombre de méthodes ont été proposées. Par exemple, [Feldman 2004, Veerara-ghavan 2008, Maitra 2009, Kimura 2011] proposent de suivre les trajectoires des abeilles dans la ruche. Leur but est d'observer et de détecter automatiquement la danse des abeilles. Il a été prouvé que cette danse est un vecteur de communication au sein de la colonie. On peut aussi mentionner [Khan 2004a] qui utilise des scènes encombrées d'intérieurs de ruches pour montrer l'efficacité de sa méthode de suivi.

Devant la ruche À notre connaissance, [Campbell 2008] est la seule référence proposant une approche de suivi d'abeilles devant la ruche. L'approche proposée permet à l'aide d'une caméra classique de compter les entrées/sorties des abeilles. Mais cela est une application indirecte issue de la collecte en 2D des trajectoires associées aux abeilles. On peut finalement citer nos propres travaux [Chiron 2013a, Chiron 2013b, Chiron 2014a] proposant diverses méthodes s'appuyant sur la stéréovision pour le suivi d'abeilles en 3D devant la ruche.

En plein air Toujours dans le contexte de l'étude des effets des pesticides, les auteurs de [Bender 2003] présentent un dispositif de télédétection par laser (ou LIDAR) permettant

7. Apilab est un bureau d'études spécialiste de la biosurveillance de l'environnement grâce à l'abeille. 10 rue Henri Bessemer, 17140 Lagord, France. (<http://www.apilab.fr/>)

de suivre un essaim d'abeilles à une distance de plus de 1 km du récepteur. Les auteurs de [Riley 1996] proposent une méthode utilisant un radar harmonique permettant de suivre la trajectoire d'une seule abeille (occasionnellement quelques abeilles) à la fois sur de longues distances.

On peut également mentionner [Estivill-Castro 2003] qui propose une approche très basique dont l'objectif est de suivre en 3D une abeille butinant une fleur à l'aide d'une caméra. Mais selon nous, les résultats ne sont pas convaincants.

4.2.3 Autres approches

Des prototypes de ruches connectées nées du projet OpenBeeLab⁸ permettent aujourd'hui de suivre en temps réel divers indicateurs tels que le poids de la colonie, la température et l'humidité à l'intérieur de la ruche.

Les travaux menés autour de l'abeille domestique au Laboratoire L3i s'intègrent dans un écosystème plus large au sein du projet Preserve. Le projet Preserve vise à organiser, croiser et mettre à disposition des chercheurs et des décideurs des données environnementales pertinentes. La plateforme⁹ mise en ligne en 2012, permet la consultation des données apicoles (statistiques d'entrées/sorties) issues de plusieurs ruchers dispersés sur le territoire français (e.g. Périgny, Bandol). Les données sont collectées via des compteurs d'abeilles aujourd'hui commercialisés par l'entreprise ApiLab. La figure 4.1 illustre le principe de partage des données qu'offre la plateforme Preserve.

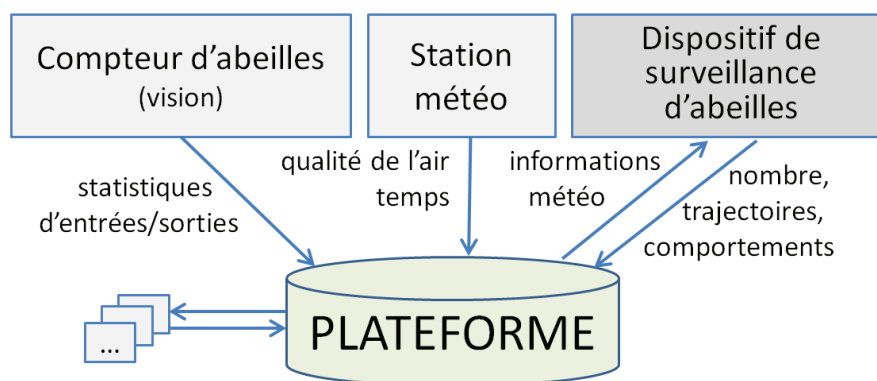


Figure 4.1 – Principe de partage des données de la plateforme Preserve.

4.3 Protocole expérimental

Cette section présente le protocole expérimental de test que nous avons mis en place en collaboration avec l'INRA Magneraud. Le but étant ici de répondre aux attentes méthodologiques du projet EPERAS. Nous verrons également comment, dans le cadre de ce protocole

8. Projet alliant chercheurs et apiculteurs dans le but de fournir sous licences libres et open-source des systèmes permettant d'alimenter des bases de données. <http://openbeelab.org/>

9. Plateforme d'agrégation, de traitement et de partage de données environnementales.

expérimental, nous avons pu anticiper un certain nombre de tests préliminaires de manière à valider certains aspects méthodologiques du projet RISQAPI.

Projet EPERAS Pour rappel, un des objectifs du projet est d'étudier les effets de produits phytosanitaires sur le comportement des abeilles en observant leurs trajectoires de vol devant la ruche. Nous le verrons par la suite, cette expérience nécessite une identification individuelle des abeilles par RFID. En effet, à l'issue d'une mise en correspondance des données d'identification RFID et des trajectoires observées par notre dispositif stéréovision, il devient alors possible de retrouver la classe d'intoxication des abeilles en vol observées par la caméra. Cette mise en correspondance permet de compléter les histoires de vie recueillies par les dispositifs RFID par des informations comportementales calculées à partir des trajectoires de vol. Cela offre des possibilités intéressantes en termes d'analyse pour les biologistes. Les méthodes nécessaires à l'accomplissement de cette expérience sont toutes décrites dans ce chapitre. Notons que la dernière brique permettant la fusion données RFID / données de trajectoires est aujourd'hui encore en cours d'implémentation.

Projet RISQAPI Nous mettons également à profit l'organisation mise en place dans le contexte du projet EPERAS pour collecter, à la fin de chaque journée d'intervention, une séquence vidéo supplémentaire. Ces séquences supplémentaires sont réservées à des fins expérimentales dans le cadre de tests préliminaires pour le projet RISQAPI. Rappelons-le, un de nos rôles au sein du projet RISQAPI est d'étudier via les comportements de vol, l'influence de la prédation par le frelon asiatique sur la diminution constatée des colonies d'abeilles. Ainsi en guise de tests préliminaires, nous avons simulé en fins de journées des perturbations de la colonie dans l'optique de détecter des anomalies (ou ruptures) au niveau des résultats de modélisation comportementale à partir des trajectoires. Cela constitue une première vérité terrain, avec la réserve que ces perturbations aient été effectuées de manière heuristique sans valeur biologique vraiment prouvée. Les perturbations en questions correspondent par exemple à : l'immobilisation d'une personne devant la ruche occasionnant une modification forcée des trajectoires d'entrées/sorties, la réalisation de mouvements brusques répétés, ou l'envoi d'un jet d'air en direction de la planche d'envol.

En complément, dans le cadre intégral du projet RISQAPI, nous avons également effectuée une campagne d'acquisition vidéo dédiée exclusivement à l'étude de l'impact du frelon. Celle-ci fut effectuée durant l'automne 2013 dans la zone atelier de la Plaine et Val de Sèvre. À l'issue de cette campagne, nous avons collecté des séquences vidéo (plus de deux heures cumulées) contenant des attaques de frelons asiatiques devant des ruches. Les données capturées durant cette campagne sont encore à l'étude, ainsi les résultats correspondant à ces séquences ne figurent pas dans ce manuscrit.

4.3.1 Contraintes de terrain

Les premiers tests sur le terrain pour la collecte de séquences vidéo à l'entrée des ruches ont été effectués en 2012 dans deux contextes différents : 1) dans le rucher de Perigny accompagné de la société ApiLab, 2) au domaine du Magneraud accompagné de biologistes de l'INRA. Ces tests nous ont enseigné que la collecte de données vidéo dans des ruchers en milieu naturel nécessite une certaine organisation, à savoir qu'il faut prévoir :

- des protections d'apiculteur (gants en cuir, vareuse avec chapeau et voile), ne facilitant pas les manipulations du matériel informatique ;
- une source électrique faible permettant d'alimenter l'ordinateur, la caméra et un support de stockage externe. Nous avons opté pour un convertisseur d'alimentation 12V/220V pour allume cigare et des rallonges électriques ;
- un accès véhicule proche des ruches (de préférence à moins de 50m) ;
- un lieu de repli pour protéger le matériel en cas d'intempéries.

De manière à ce que la méthodologie proposée soit reproductible dans le cadre d'études réelles, nous nous sommes imposés une certaine rigueur expérimentale conformément aux bonnes pratiques des biologistes. Cela a rendu les manipulations terrains particulièrement complexes. Notamment en ce qui concerne 1) la quantité de données recueillies s'élevant à plus de 20 heures de vidéo cumulées (correspondant à des centaines de milliers de trajectoires), et 2) le calendrier des interventions s'étendant sur plus d'un mois pour couvrir l'évolution biologique de la colonie dans le temps.

4.3.2 Calendrier des interventions

Concernant le calendrier planifié dans le cadre du projet EPERAS, la phase d'acquisition des séquences vidéo s'est étendue sur la période estivale de mi-juin à fin juillet 2013. La figure 4.2 illustre le calendrier effectif des interventions effectuées par le L3i et l'INRA Magneraud. Le L3i fut chargé d'acquérir des séquences vidéo des abeilles en vol à l'entrée. L'INRA fut chargé d'introduire chaque semaine, dans la colonie, de nouveaux lots d'abeilles (appelés cohortes) exposées volontairement à divers niveaux d'intoxication. De plus, les abeilles sont équipées à la naissance d'une puce RFID, ce qui permet de suivre leur histoire de vie (horodatage des entrées/sorties de la ruche) jusqu'à leur mort. Les abeilles prennent leur premier envol autour de la troisième semaine. C'est à partir de ce moment, que nous avons commencé à recueillir des données vidéo. Les acquisitions ont été planifiées l'après-midi entre 14h et 18h30, ce qui correspond à la période de pleine activité de la colonie lorsque le ciel est dégagé. Nous avons en moyenne capturé 9 séquences vidéo d'une quinzaine de minutes par jour. Telle que mentionnée précédemment, la dernière capture fut réservée pour capturer l'effet d'une perturbation sur la colonie en vue d'effectuer des tests dans le cadre du projet RISQAPI.

Les séquences vidéo de la journée du 23/08/2013 se sont révélées inexploitable suite à une erreur de manipulation sur le terrain. La journée du 25/08/2013 ne comporte, quant à elle, que 4 captures exploitables, les 4 dernières ayant été victimes d'un problème de sauvegarde. Finalement, nous comptons au total 74 séquences exploitables, pour un total d'environ 20h de vidéo RGB-D, ce qui représente plus d'1 To de données brutes à stocker et à traiter.

L'annexe C présente un exemple de fiche de suivi des acquisitions effectuées par nos soins. Cette fiche a pour objectif de répertorier des informations supplémentaires (e.g. météo, activité, événements marquants) permettant par la suite, à l'image d'une vérité terrain, de retrouver certaines tendances potentiellement observables sur les résultats issus de l'analyse comportementale à partir des trajectoires.

	LUNDI	MARDI	MERCREDI	JEUDI	VENDREDI
JUN 2013	<u>10</u>	<u>11</u> Intro Cohorte 1 (Fabrice)	<u>12</u>	<u>13</u>	<u>14</u>
	<u>17</u>	<u>18</u> Intro Cohorte 2 (Fabrice)	<u>19</u>	<u>20</u>	<u>21</u> Test / Calibration (Fab., Gui., Bruno)
	<u>24</u>	<u>25</u> Intro Co.3 (Fabrice)	<u>26</u> Acquisition 8 captures (Guillaume)	<u>27</u> Acquisition 9 captures (Guillaume)	<u>28</u>
JUILLET 2013	<u>1</u>	<u>2</u> Intro Co.4 (Fabrice)	<u>3</u> Acquisition 9 captures (Guillaume)	<u>4</u>	<u>5</u> Acquisition 9 captures (Bruno)
	<u>8</u>	<u>9</u> Intro Co.5 (Fabrice)	<u>10</u> Acquisition 9 captures (Guillaume)	<u>11</u>	<u>12</u> Acquisition 9 captures (Bruno)
	<u>15</u>	<u>16</u> Intro Cohorte 6 (Fabrice)	<u>17</u> Acquisition 9 captures (Bruno)	<u>18</u>	<u>19</u> Acquisition 9 captures (Guillaume)
	<u>22</u>	<u>23</u> Intro Co.7 (Fabrice)	<u>24</u> Acquisition 9 captures (Bruno)	<u>25</u> Acquisition 4 captures (Guillaume)	<u>26</u>

 Manipulations L3i (Informatique)
 Manipulations INRA (biologistes)

Figure 4.2 – Calendrier effectif des interventions 2013 conjointes L3i/INRA, principalement pour l'étude toxicologique liée à l'abeille domestique prévue dans le cadre du projet EPERAS.

4.3.3 Dispositif d'acquisition

La figure 4.3 présente le dispositif d'acquisition stéréovision utilisé lors de la campagne d'acquisition vidéo de l'été 2013. L'ensemble caméra stéréovision/trépied est fixé sur un toit de ruche amovible. Ce toit amovible est positionné sur la ruche à étudier au début de chaque série (généralement un après-midi) d'acquisition. Ainsi, le dispositif demeure stable, ce qui a permis de capturer toutes les séquences vidéo de la saison au sein d'un repère spatial commun. Dans le but d'améliorer la qualité de l'appariement lors de la mise en correspondance stéréo, nous avons moucheté la planche d'envol avec des marques noires. Nous le verrons plus tard, cela a malheureusement complexifié le problème de détection des puces RFID sur les images (cf. Section 4.5.2).

Tout au long de la saison, en parallèle des captures vidéos et de manière récurrente, l'INRA a recueilli des données d'entrées/sorties des abeilles issues des différentes cohortes intégrées à la colonie lors des phases d'introduction. Les abeilles en question ont été équipées d'un tag RFID à leur naissance. Cette collecte de données s'est effectuée aux niveau des deux rangées de 5 lecteurs RFID positionnés à l'entrée de la ruche.

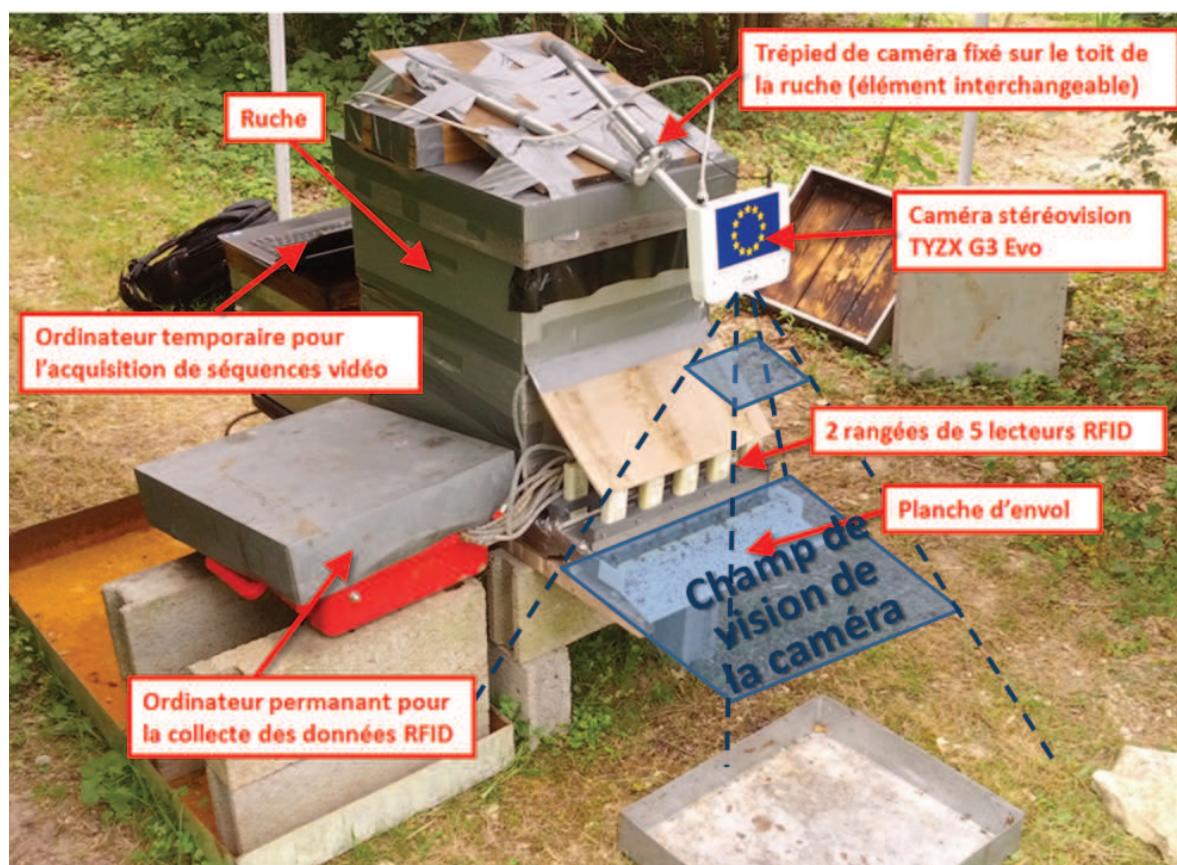


Figure 4.3 – Dispositif d'acquisition stéréovision à l'entrée de la ruche. Le trépied servant à tenir la caméra est fixé sur un toit de ruche mobile. Cela permet de conserver la même configuration (inclinaison, hauteur de la caméra) lors des déplacements. Sur des ruches standards, les séquences vidéo sont ainsi acquises dans un même référentiel. À savoir que le dispositif d'acquisition doit être mis à l'abri après chaque journée d'acquisition, permettant par la même occasion aux apiculteurs d'intervenir sur la ruche entre 2 journées de manipulation.

4.3.4 Préparation des données et prétraitements

Stockage des données La caméra stéréovision G3 Evo de notre dispositif enregistre les séquences RGB-D et les métadonnées dans un format propriétaire volumineux. Les métadonnées stockent entre autres, les paramètres de calibration de la caméra (utilisés pour la projection 2D/3D via l'équation 2.13), ainsi que l'horodatage (ou "timestamps" en anglais) associé à chaque image. Les timestamps sont exprimés relativement à l'horloge interne de la caméra. Les timestamps sont intéressants pour restituer la séquence vidéo dans le temps, mais ils sont surtout essentiels pour les phases de prédictions/corrections effectuées lors du suivi de cibles. En effet, la fréquence d'acquisition n'est pas stable. On note un écart de temps entre deux images successives allant du simple (19 ms) au triple (57 ms).

Pour des raisons pratiques de stockage, il est préférable d'extraire et de recompresser ces données, originellement sous format propriétaire, dans un format plus facilement manipulable. Les séquences d'images d'intensité, et de cartes de profondeur ont respectivement été encodées sur 8 bit et 16 bit au format Motion JPEG2000 (sans perte de qualité). Puis, les métadonnées

ont été converties au format texte.

Finalement, les 74 séquences vidéo correspondent à un total cumulé de 2 976 678 couples d'images d'intensité/carte de profondeur définis à une résolution de 740×468 px. Cette phase de ré-encodage a permis de gagner un facteur 1.6 au niveau du stockage (descendant le poids total des données à 750 Go). Et à notre avantage, le format Motion JPEG 2000 offre une grande flexibilité pour la manipulation, notamment parce qu'il est compatible avec des lecteurs multimédia grand public (e.g. VLC). La phase d'extraction de données et de ré-encodage a duré environ une centaine d'heures sur notre machine équipée d'un Core i5 cadencé à 2.5 GHz.

Synchronisation des horloges Le dispositif mis en place pour notre expérience (cf. Figure 4.3) collecte des données temporelles relativement à 3 horloges différentes :

- l'horloge interne de la caméra,
- l'horloge principale du PC également utilisée pour horodater les notes manuscrites (cf. Annexe C),
- l'horloge du dispositif de capture RFID.

Les différences entre les horloges $\Delta_{\text{Cam_PC}}$ et $\Delta_{\text{PC_RFID}}$ (cf. Figure 4.4) peuvent changer d'une capture sur l'autre. En effet, l'horloge interne de la caméra est indépendante de l'horloge du PC auquel elle est connectée, et est ainsi réinitialisée à chaque branchement. De plus, l'ordinateur contrôlant le dispositif RFID est susceptible de changer entre deux manipulations. Ceci s'est d'ailleurs produit durant notre campagne de capture. Pour permettre une fusion ultérieure des données, il est impératif de noter rigoureusement, au début de chaque capture, des points de synchronisation (différences d'horloges). Cela permet de recaler toutes les données acquises au sein d'un même repère temporel, que nous avons fixé comme étant, celui de l'horloge du PC principal.

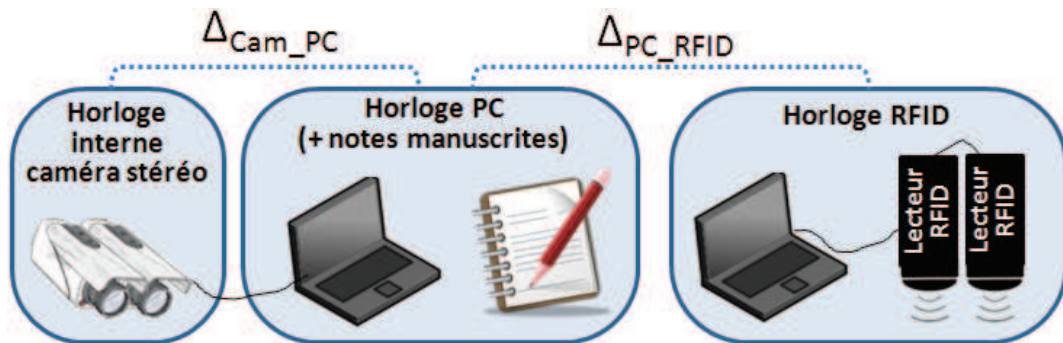


Figure 4.4 – Données acquises selon trois horloges différentes.

Renseignement des coordonnées de la planche d'envol Il est nécessaire pour chaque configuration de scène (position de la caméra par rapport à la planche d'envol), de renseigner le polygone 2D englobant la planche d'envol (cf. Annexe A.3). Dans le cadre de notre expérience, la configuration du dispositif est la même pour toutes nos captures, ainsi l'opération de renseignement des coordonnées peut être effectuée une fois pour toutes. Grâce aux coordonnées 2D englobant la planche d'envol, on peut calculer le maillage 3D de la surface (cf. Section 2.4.3). Outre le fait que cela permet d'offrir un affichage plus convivial des résultats

de segmentation et suivi de cibles, cela permet surtout de calculer la distance entre les cibles et la planche d'envol dans le repère 3D. Ce qui est notamment un prérequis pour la mise en application de notre contribution proposée dans la section 2.3.2. Cela permet également de déduire la nature des traces (i.e. départ, arrivée, simple passage dans le champ de vision), ce qui s'avéra utile par la suite, par exemple lors des phases d'analyse comportementale.

4.4 Traitement des séquences vidéo

Cette section détaille pas à pas l'application des différentes étapes de notre chaîne de traitement présentée dans les chapitres 1, 2 et 3.

4.4.1 Détection des abeilles par segmentation HIDS

En premier lieu, nous appliquons notre méthode de segmentation HIDS sur l'ensemble des 74 séquences vidéo capturées. À l'issue de ce traitement, l'ensemble des vidéos pesant au total 765 Go (en format Motion JPEG2000) sont devenues 30 436 022 positions 3D (ou 2D lorsque la profondeur est indéfinie) accompagnées de quelques métadonnées associées (timestamps et ellipses englobantes), pour un volume total de 4.5 Go.

Cette phase de détection de cibles dure en moyenne 0.3 sec/image sur notre machine équipée d'un Core i5 cadencé à 2.5 GHz, soit un total de 10 jours pour l'ensemble des 74 séquences capturées. L'implémentation de la méthode est faite en Matlab et est relativement peu optimisée. Malgré une fréquence d'acquisition élevée (allant jusqu'à 53 Hz), et donc un nombre d'images RGB-D important à traiter par seconde, nous pensons qu'il est possible de faire fonctionner notre algorithme de segmentation en temps réel, moyennant un effort plus important d'implémentation.

La figure 4.5 montre, sous forme 2D (couple intensité/profondeur) et 3D, un exemple d'abeilles en vol détectées via notre méthode de segmentation HIDS en condition de forte activité. Dans cet exemple, 26 abeilles en vol ont été simultanément détectées, dont deux avec une profondeur indéfinie. On note aucune détection manquée, mais d'éventuelles fausses détections au niveau de la sortie de la ruche. Cette zone à forte activité est plutôt sensible, mais connaissant le modèle 3D de la planche d'envol, il est facile de filtrer ces mauvaises détections en fixant un seuil de distance par rapport à la planche.

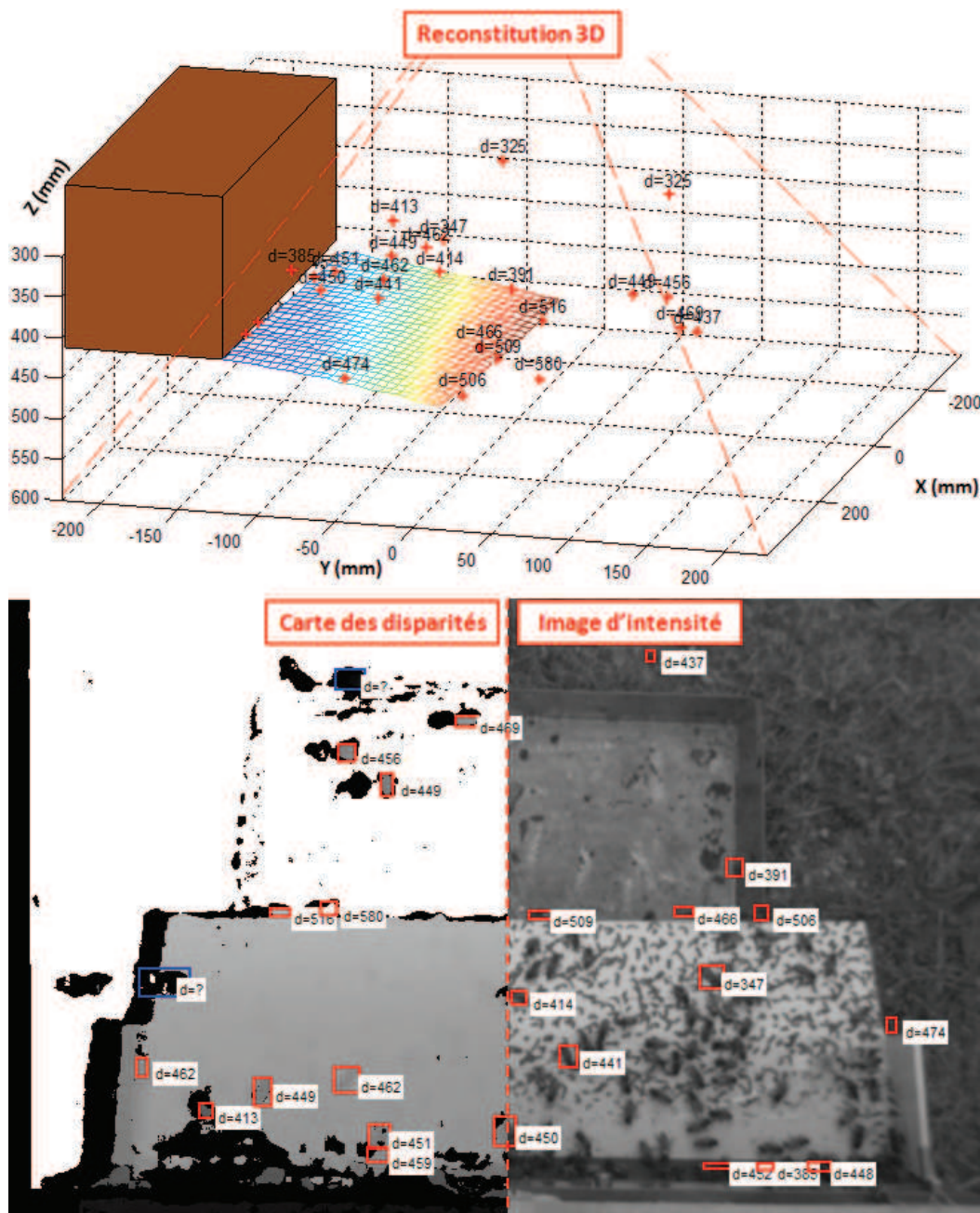


Figure 4.5 – Exemple de détection d'abeilles en vol via notre méthode de segmentation HIDS en condition de forte activité (capture du 23/07/2013 à 17h00). La reconstitution 3D de la scène est effectuée automatiquement grâce aux données de segmentation et au maillage de la planche d'envol (cf. méthode présentée en section 2.4.3). Les étiquettes associées aux cibles mentionnent la profondeur calculée. Sur le modèle 3D, les pointillés rouges indiquent la limite du champ de vision de la caméra. Les abeilles, dont la profondeur peut être calculée, sont marquées en rouge, tandis que les abeilles, dont la profondeur ne peut pas être retrouvée, sont marquées en bleu. Sur ce couple d'images intensité/profondeur, 26 abeilles en vol ont été détectées, dont 2 sans information de profondeur, par conséquent non positionnables dans l'espace 3D.

4.4.2 Calcul des trajectoires par suivi multi-cibles

Nous appliquons sur les données de localisation obtenues à l'étape précédente, notre méthode de calcul de trajectoires en nous appuyant sur le filtre de Kalman et sur GNN (cf. Section 2.3.1). À l'issue de ce traitement, l'ensemble des 30 436 022 positions 3D d'abeilles détectées deviennent 861 337 trajectoires 3D réduisant le poids des données à environ 1.6 Go (incluant des métadonnées comme des informations sur la vitesse 3D et les horodatages associés).

Cette phase de calcul des trajectoires nécessite environ 5 jours. De par la nature de l'algorithme de suivi, les calculs sont bien ralentis lorsque que le nombre de cibles à traiter devient élevé (e.g. plus de 20 cibles simultanées). Sans ces situations ponctuelles d'encombrement, les calculs des trajectoires seraient bien plus rapides. Dans ce cadre, on pourrait imaginer traiter les morceaux de séquences très encombrées en parallélisant les calculs par exemple.

La figure 4.6 illustre un exemple de calcul de trajectoires d'abeilles en vol en condition de forte activité. Cet exemple montre 12 traces simultanées et confirmées d'abeilles en vol devant la ruche. Les fausses détections liées aux cibles situées autour de la planche d'envol ont été en grande partie supprimées en appliquant un seuil nécessaire de 2 cm entre la planche d'envol et une abeille en vol.

4.4.3 Extraction des caractéristiques

La troisième phase concerne le calcul des signatures à partir des trajectoires 3D. Nous avons constaté un grand nombre de trajectoires naissantes non-abouties dues notamment à des fausses détections. Des abeilles grimpent parfois les unes sur les autres et dépassent ainsi temporairement la limite des 2 cm de distance par rapport à la planche d'envol, ce qui engendre de courtes trajectoires à proximité de la planche. Ce phénomène peut d'ailleurs être amplifié si le maillage 3D de la surface de la planche d'envol n'est pas précis. Nous proposons de filtrer ces artefacts pour ne conserver que les traces intéressantes. Voici les caractéristiques des traces que nous choisissons de préserver :

- traces construites sur plus de 10 observations ;
- traces dont la longueur totale mesure plus de 10 cm ;
- traces dont l'espace entre deux positions d'abeilles ne dépasse pas les 15 cm.

Ainsi, 315 154, soit 36% des trajectoires, sont conservées sur l'ensemble des 861 337 trajectoires originalement renvoyées par la méthode de suivi de cibles.

Les 10 caractéristiques extraites à partir de chaque trajectoire sont celles mentionnées dans la section 3.3.3.2. Cette étape de filtrage et de calcul des signatures n'a duré qu'une trentaine de minutes sur notre machine équipée d'un Core i5 cadencé à 2.5 GHz. La figure 4.7 illustre pour un exemple de capture, une projection de l'ensemble des signatures dans un espace 3D (via une ACP).

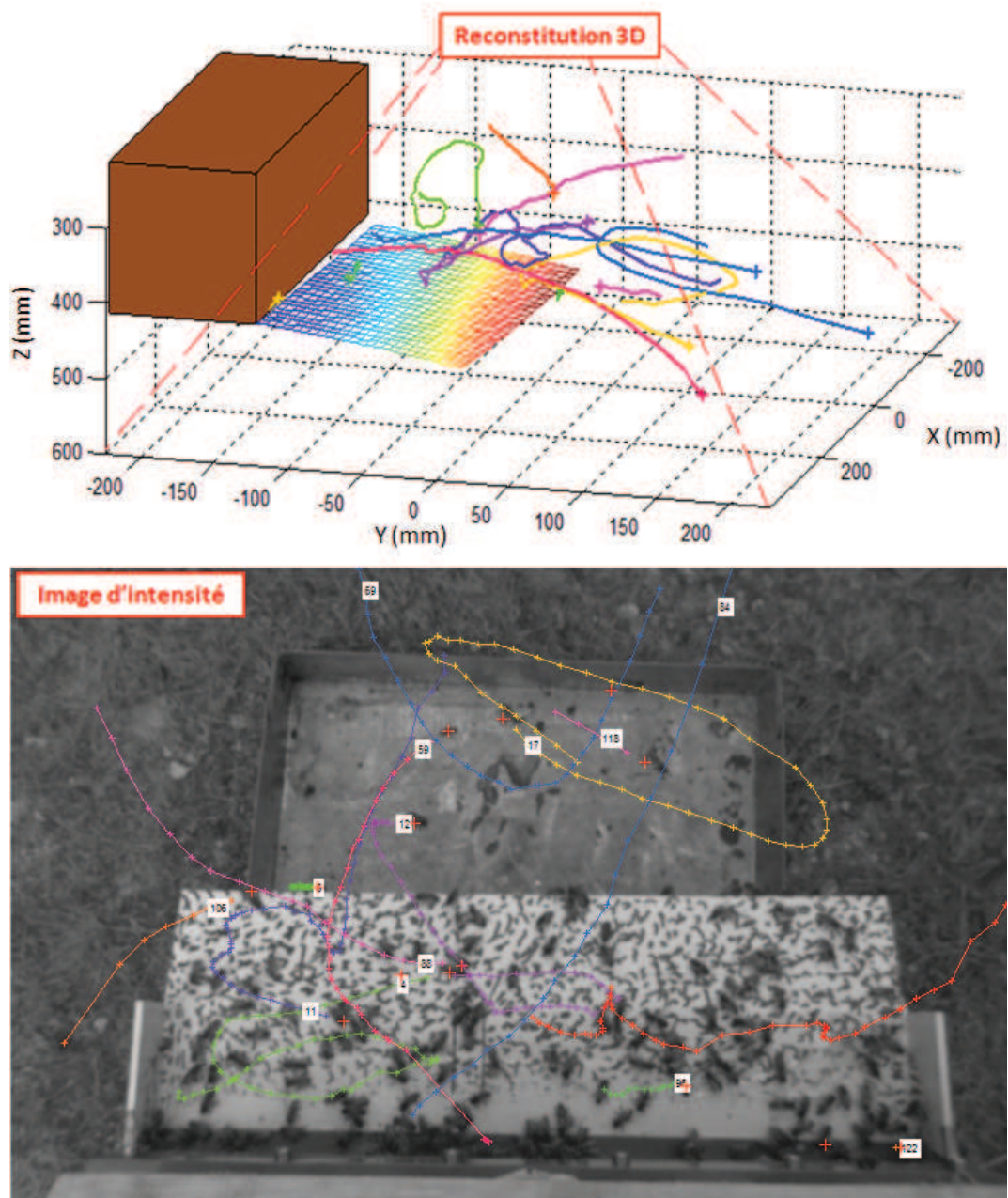


Figure 4.6 – Exemple de calcul de trajectoires d'abeilles en vol via la méthode proposée en section 2.3.1 en condition de forte activité (capture du 23/07/2013 vers 17h00). Les traces naissantes (depuis moins de 3 itérations) ne sont pas représentées ici.

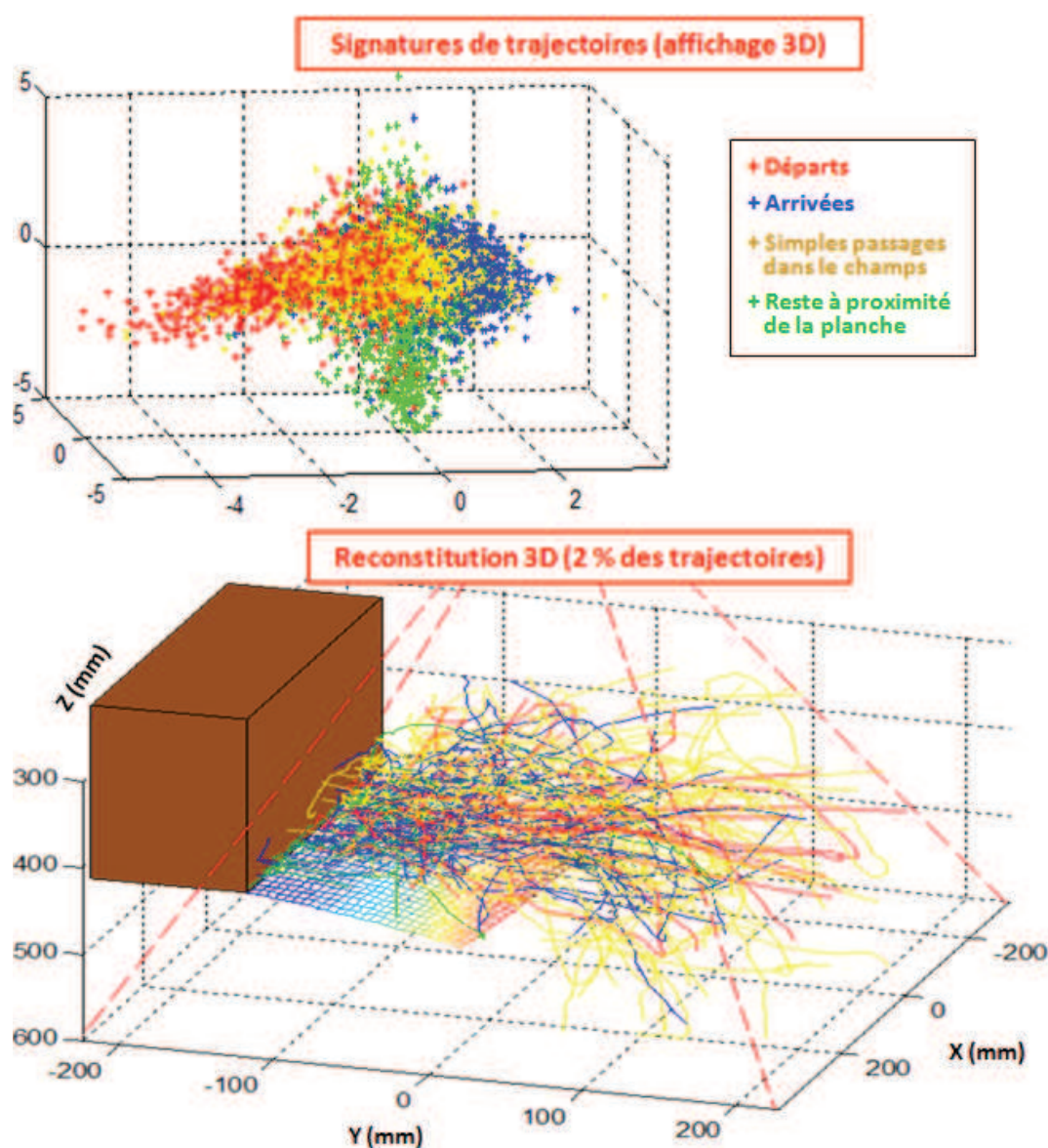


Figure 4.7 – Projection dans un espace 3D (via une ACP) des signatures de l'ensemble des trajectoires d'une capture (le 23/07/2013 de 17h00 à 17h15). Le reconstitution 3D (en bas) illustre les trajectoires ainsi que leur nature codée sur 4 couleurs. Pour éviter une surcharge de la figure, cette reconstitution se limite à un échantillon de 170 traces, soit 2% de l'ensemble des trajectoires de la capture qui en compte en tout 8485.

4.4.4 Modélisation MLC-HDP pour RISQAPI

À ce stade de l'expérience, nous n'avons pas fait le lien avec les données RFID. Ainsi, les résultats qui vont suivre sont plutôt à replacer dans le contexte du projet RISQAPI. En effet, dans ce projet, la connaissance des conditions expérimentales (e.g. classe d'intoxication) des abeilles individuelles n'est *a priori* pas nécessaire.

Cette section détaille l'approche choisie pour structurer, selon une hiérarchie en *captures* et *sousclips*, l'ensemble des 315 154 signatures de trajectoires. Ensuite, à l'issue de la mise en œuvre du MLC-HDP (cf. Chapitre 3), nous présentons les résultats de clustering selon les trois niveaux (*atome d'activité*, *activité* et *comportement*).

4.4.4.1 Structuration du jeu de données selon trois niveaux hiérarchiques

Dans le projet RISQAPI, nous cherchons à découvrir et à modéliser des comportements types adoptés par une ou plusieurs colonies d'abeilles en fonction du temps. Le but étant de pouvoir effectuer des comparaisons de comportements à différentes échelles temporelles :

- **échelle courte** (*sousclip* d'une minute) : durée pendant laquelle la colonie va être affectée par un événement soudain (e.g. réponse à l'attaque d'un frelon asiatique, essaimage¹⁰) ;
- **échelle longue** (*capture* d'1/4 d'heure) : durée permettant de capturer des comportements de fond (e.g. stress dû à la présence prolongée du frelon asiatique, effet météo, maladie, intoxication).

En regroupant ensemble les *captures* similaires, ainsi que les *sousclips* similaires, il devient alors possible d'identifier des périodes de stabilité, des ruptures, ou même des phénomènes cycliques récurrents au sein des *comportements* ou *activités* de la colonie.

Ainsi, l'idée est de structurer le jeu de données sous une forme hiérarchique de manière à ce que chaque entité "*capture*" soit associée à un des *comportements* trouvés. La figure 4.8 représente la structure à trois niveaux que nous allons utiliser par la suite.

Les résultats présentés par la suite ont été obtenus à l'issue de 46 itérations, lorsque la log-vraisemblance négative (cf. Section 3.3.4) atteint un minimum. Les temps de calculs sont d'environ une heure sur notre machine.

10. Départ soudain d'une partie des abeilles de la ruche pour former une nouvelle colonie.

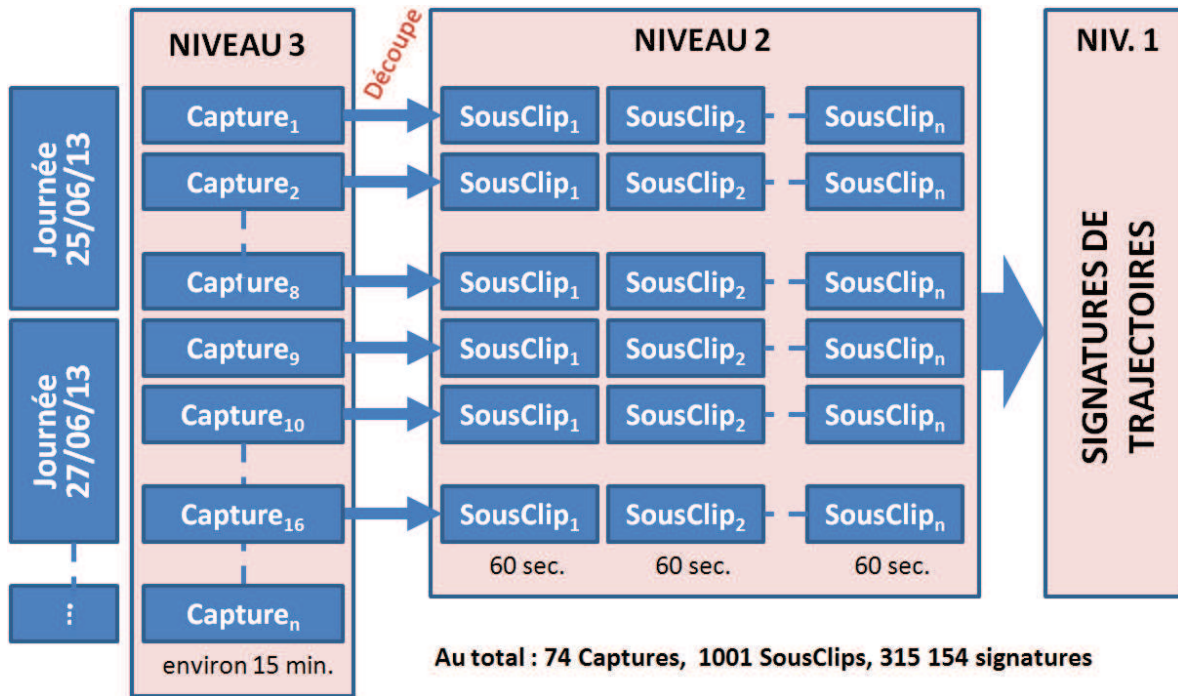


Figure 4.8 – Structuration des signatures de trajectoires au sein d'un jeu de données hiérarchisé. Au total, on compte 315 154 signatures réparties au sein de 1001 *sousclips* issus de 74 *captures* vidéo.

4.4.4.2 Résultats Niv.1 : Atomes d'activité

Les *atomes d'activité* sont les clusters de base définis par des paramètres ϕ_k (tirés de la distribution de base H). La figure 4.9 montre comment le MLC-HDP a regroupé l'ensemble des signatures de trajectoires en 12 clusters de type "*atome d'activité*". Le modèle de mélange de gaussiennes multi-variées semble adapté pour modéliser notre ensemble de données. En termes d'interprétations, on peut fonder des hypothèses en s'appuyant sur la structure et les notions de "départs/arrivées/..." illustrées dans la figure 4.7. En effet, on retrouve une structure analogue dans la figure 4.9, et les clusters modélisent bien ces différentes notions. Par exemple, on peut deviner que les 2 clusters les plus à gauche sur la figure, englobent une grande partie des trajectoires au départ de la ruche. Ces deux clusters correspondent vraisemblablement à 2 types de trajectoires (e.g. départs directs de butineuses, ou départs plus chaotiques correspondant à des vols d'apprentissage.)

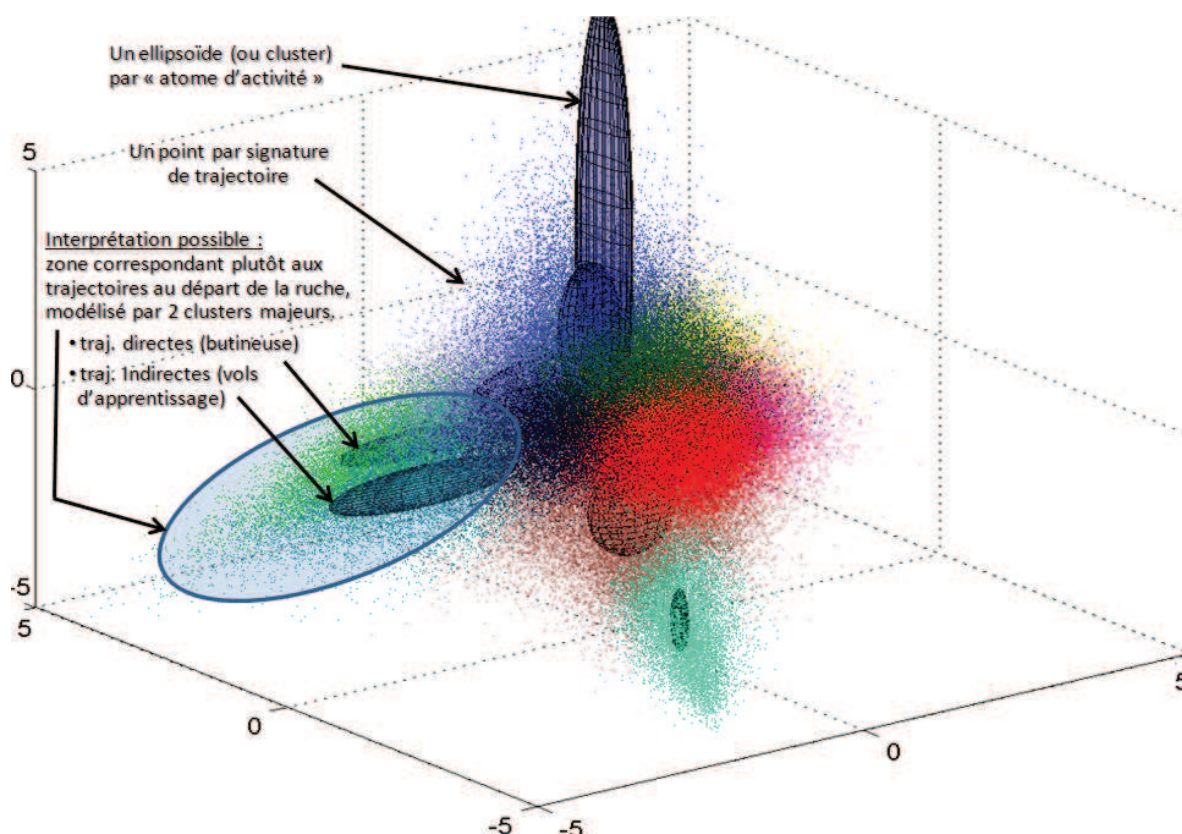


Figure 4.9 – Résultat du clustering MLC-HDP (niv. 1 - *atome d'activité*), sur l'ensemble des données de trajectoires, couvrant la totalité des *captures* acquises durant l'été 2013 à l'INRA Magneraud. Chaque cluster est représenté par un ellipsoïde, et chaque trajectoire est représentée par un point dont la couleur représente son cluster d'affectation. Les interprétations associées aux zones et aux clusters s'appuient sur les notions de "départs/arrivées..." données dans la figure 4.7, mais ces dernières restent purement hypothétiques. Ces éléments sont affichés dans un espace 3D issue d'une ACP sur l'espace des caractéristiques originales.

4.4.4.3 Résultats Niv.2 : Activités

Chaque *activité* est une distribution, que l'on peut interpréter par un modèle d'ensemble de poids sur les *atomes d'activité*. Dans notre modèle, Chaque *sousclip* est associé à une *activité* donnée. Sur notre exemple, le MLC-HDP découvre 12 modèles d'ensembles de poids. Ces 12 *activités* distinctes servent à couvrir les diverses situations rencontrées au sein des 1001 *sousclips*.

La figure 4.10 montre les *activités* associées aux 21 derniers *sousclips* de la journée du 19/07/2013, ce qui correspond aux 21 minutes capturées à partir de 17h30. Durant les 15 premières minutes, l'activité de la colonie n'est pas influencée par un facteur extérieur. Puis, pendant les 3 minutes suivantes, des perturbations importantes ont été provoquées. À savoir, des mouvements très brusques à proximité de la ruche. Ces mouvements ont tendance à énerver les abeilles et à potentiellement modifier leurs couloirs de vol. Enfin, les 3 dernières minutes correspondent à un éventuel retour à la normal sans aucune perturbation. Sur l'exemple présenté sur la figure 4.10, on observe une réponse à la perturbation sous forme d'une rupture

à la 15^{ème} minute au niveau des *activités* associées aux *sousclips*.

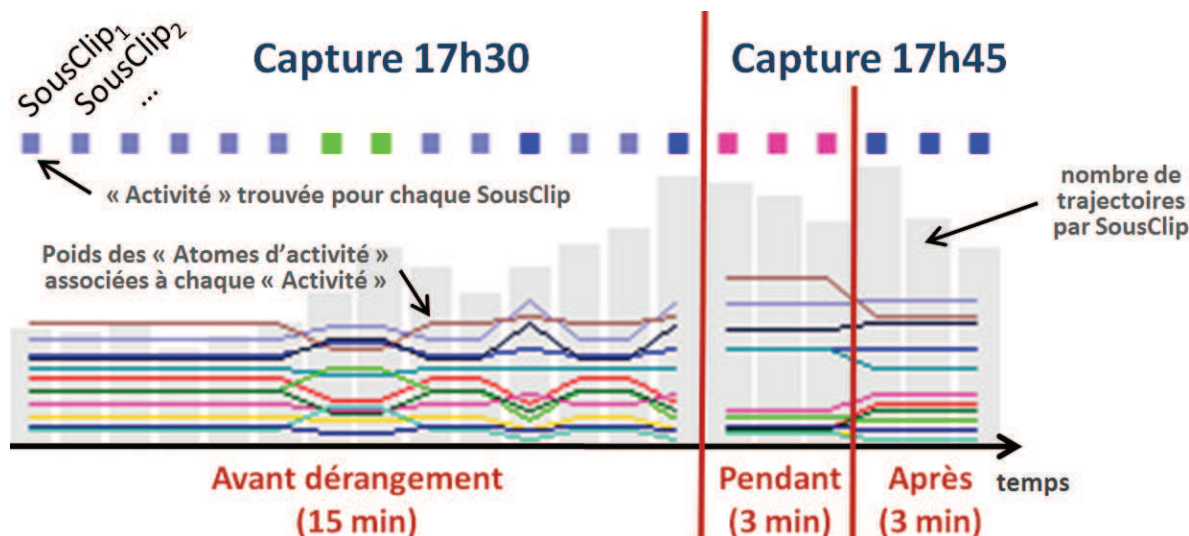


Figure 4.10 – Résultat de clustering MLC-HDP (niv. 2 - "activité") sur les 21 derniers *sousclips* de la journée du 19/07/2013. Les 6 dernières minutes de la journée d'acquisition sont réservées pour tester l'effet d'une perturbation provoquée sur la colonie.

4.4.4.4 Résultats Niv.3 : Comportements

Tout comme les *activités* sont des modèles d'expression des *atomes d'activité*, les *comportements* sont des modèles d'expression d'*activités*. Dans notre modèle, chaque *capture* de 15 minutes est associée à un *comportement* donné. Sur notre exemple, le MLC-HDP trouve 10 *comportements*. Ces 10 *comportements* distincts couvrent les diverses situations rencontrées au sein des 74 *captures*.

La figure 4.11 montre les résultats de la modélisation MLC-HDP au niveau des *comportements* sur les 74 *captures*. Le premier constat est que l'on observe une certaine stabilité comportementale au cours du temps. Notamment, on retrouve peu de *comportements* qui se distinguent dans le temps. Parfois même, les *comportements* de débuts d'après-midi sont similaires d'une journée sur l'autre (e.g. du 2/07/2013 au 12/07/2013). Deuxième constat, les simulations de perturbation effectuées en fins de journées ont été clairement détectées sur deux *captures* sur sept. Dans les cinq autres cas, les *activités* associées aux *sousclips* restent stables. Cela peut s'expliquer par le fait que ces perturbations ont possiblement affecté un nombre d'abeilles trop faible. De plus, le comportement de la colonie dépend sans aucun doute d'une multitude de facteurs parmi lesquels nos perturbations ont joué dans la majorité des cas un rôle mineur. Il faut savoir également que ces expériences ont été dispensées tour à tour par deux personnes (cf. Figure 4.2) et possiblement de manières différentes. Finalement, en fonction de l'activité de la ruche (au sens "nombre d'abeilles"), les effets sont probablement plus ou moins perceptibles. Au vu de ces résultats sur des données réelles, on peut dire que notre approche est efficace pour détecter des périodes stationnaires, en revanche, nous n'avons pas d'indice fiable quant à son efficacité concernant la détection de transitions.

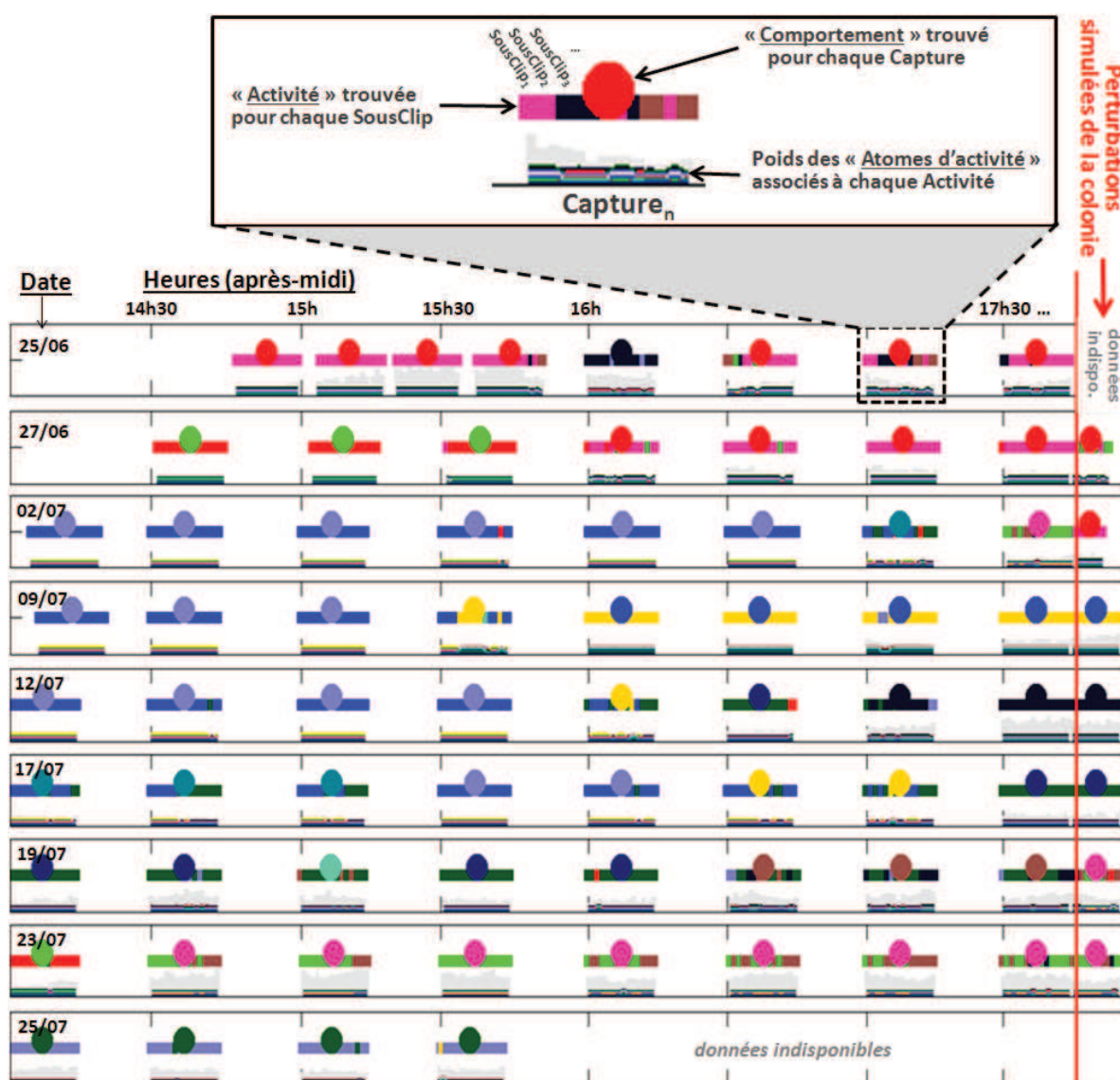


Figure 4.11 – Résultat de la modélisation MLC-HDP (niv. 3 - "comportements") sur l'ensemble des 74 captures acquises durant l'été 2013 à l'INRA Magneraud. En complément, à la fin de chaque journée de manipulation, à partir de 17h45, des perturbations de la colonie ont été réalisées (e.g. positionnement devant la ruche, mouvements brusques). Une capture est détaillée dans la partie supérieure de la figure, sur laquelle on retrouve les mêmes éléments que dans la figure 4.10. Ici, ce ne sont pas les détails au niveau des sousclips qui sont intéressants, mais bien le comportement associé à chaque capture.

4.4.5 Synthèse sur les performances

Cette section propose, via la figure 4.12, une synthèse des temps de calcul constatés au niveau des différentes étapes de la chaîne. Les expérimentations ont été effectuées sur notre machine équipée d'un Core i5 cadencé à 2.5 GHz. Le code est écrit majoritairement en matlab, et relativement peu optimisé compte tenu de la complexité des algorithmes. Le tableau 4.1

regroupe pour chaque étape de la chaîne de traitement, les principaux facteurs et paramètres à manipuler.

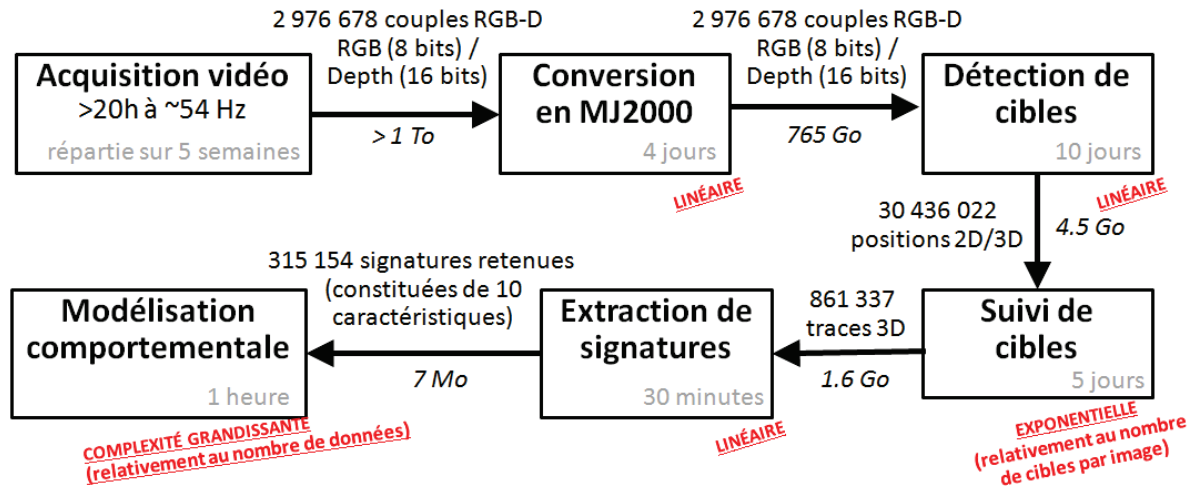


Figure 4.12 – Synthèse des performances (en termes de temps de calcul) constatés lors de l'application des différentes étapes de la chaîne. Les temps indiqués sont exprimés de manière globale, car cela est plus pertinent que des moyennes qui sont dans ce cas, sujettes à des écarts-types importants.

Étape	Principaux facteurs
Acquisition vidéo	distance d'observation raisonnable, fréquence élevée, résolution adaptée à la taille des cibles (> 8px par cible).
Détection de cibles	tailles des éléments structurants adaptées à la taille et à la nature des cibles.
Suivi de cibles	nombre de meilleures hypothèses H et profondeur de l'arbre P donnant lieu à un compromis entre la performance de la méthode et le temps de calcul.
Extraction de signatures	pertinence des caractéristiques plus importante que la quantité.
Modélisation comportementale	hyperparamètres α influençant le nombre de cluster et γ qui selon sa valeur initiale favorise ou non la convergence du modèle.

Tableau 4.1 – Principaux facteurs à retenir pour chaque étape de la chaîne de traitement.

4.5 Identification des trajectoires d'abeilles

La modélisation comportementale de la colonie en fonction du temps, telle qu'elle est présentée dans la section précédente, donne un indicateur, *capture* par *capture*, sur l'évolution du comportement de la colonie de manière globale. Cependant, dans le cadre de notre étude toxicologique du projet EPERAS, l'intérêt d'un tel modèle reste limité. En effet, il est essentiel de prendre en compte le niveau d'intoxication des abeilles pour comparer leurs *comportements*.

Le calendrier des interventions (cf. Figure 4.2) indique les dates d'introduction de 7 cohortes d'abeilles au sein de la colonie. Chaque cohorte est composée d'environ 330 abeilles. Les cohortes sont un mélange à part égale d'abeilles exposées à 3 différents niveaux sublétaux de contamination à un pesticide (2.7ppb, 4ppb et 6ppb de thiamétoxam), et d'abeilles témoins non intoxiquées. Pour rappel, les abeilles sont équipées à la naissance d'un tag RFID. Ce dernier est associé à un identifiant unique, ce qui permet de faire le lien avec l'état d'intoxication de l'abeille porteuse du tag. Au total, pas moins de 2300 abeilles taguées (sur environ 30 000 individus) ont été introduites dans la colonie dans le cadre du protocole expérimental présenté dans la section 4.3.

Les lecteurs RFID (cf. Figure 4.3) disposés à l'entrée de la ruche capturent 24h/24h les entrées/sorties des abeilles équipées d'un tag RFID. L'ensemble des entrées/sorties d'une abeille reflète son histoire de vie. Aux vues des résultats expérimentaux issus des données RFID de la campagne d'acquisition de l'été 2013, il est difficile de conclure sur l'influence ou non, sur histoire de vie des abeilles, du pesticide testé [Requier 2014]. Ainsi, afin d'apporter des données supplémentaires, nous proposons dans cette section de nous intéresser aux trajectoires en vol des abeilles taguées. Cela implique une fusion d'information qui s'avère, nous le verrons, non triviale.

4.5.1 Mise en correspondance des histoires de vie avec les trajectoires

La chaîne de traitement présentée dans les 3 premiers chapitres permet de collecter les trajectoires 3D des abeilles en vol devant la ruche. Notre approche ne permet pas de suivre les abeilles sur la planche. Or, lors de leurs arrivées, les abeilles se posent n'importe où sur la planche d'envol pour ensuite se déplacer (ou non) jusqu'à l'entrée de la ruche où se trouvent les lecteurs RFID. Lors des départs, elles adoptent le même comportement : elles se déplacent sur la planche d'envol un certain temps et une certaine distance avant de prendre leur envol (ou non). La figure 4.13 illustre le problème des données manquantes pour la mise en correspondance entre les trajectoires et les données de passages sous les lecteurs RFID.

En général, la durée entre le passage sous le lecteur RFID et l'envol ou l'atterrissage d'une abeille ne dépasse pas quelques secondes. Nous pouvons mettre en correspondance les trajectoires de type "départ" avec les séquences RFID de type "AB" signifiant une sortie de ruche (cf. Section 4.5.3). De la même façon, nous pouvons mettre en correspondance les trajectoires de type "arrivée" avec les séquences RFID "BA" signifiant une entrée dans la ruche. Ainsi, afin de restreindre le nombre de trajectoires susceptibles d'être candidates à l'association avec une séquence RFID donnée, il est indispensable de rejeter les trajectoires engendrées par des abeilles non-porteuses de puce RFID, ce qui nécessite d'identifier sur les images si une abeille est porteuse ou non d'une puce. De plus, pour éviter toutes erreurs, nous choisissons de ne garder que les situations où il n'existe qu'une seule association possible entre trajectoire visuelle et séquence RFID candidate dans une fenêtre de temps donnée.

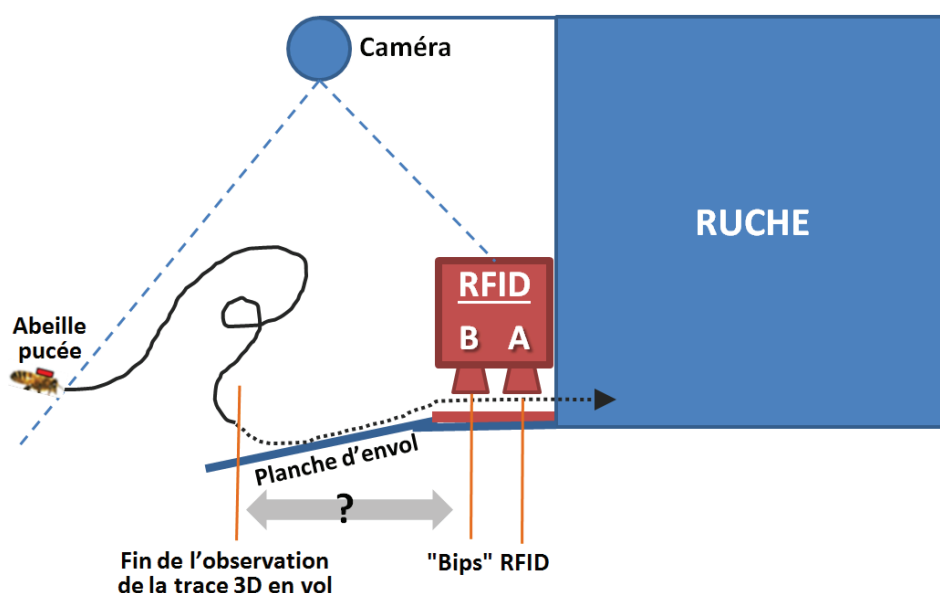


Figure 4.13 – Schéma de profil illustrant le problème de raccordement des données visuelles (disponibles uniquement pour les abeilles en vol) et les données RFID (disponibles uniquement à l'entrée de la ruche). La fréquentation de la planche d'envol et le temps moyen passé sur la planche sont à prendre en compte.

4.5.2 Détection visuelle des puces RFID

Il existe des méthodes de détection de tag sur le dos des abeilles [Hendriks 2012, Chen 2012]. Ces approches visent à identifier les abeilles via des marqueurs possédant un codage visuel. Dans notre cas, les tags sont uniformes, et contrairement aux travaux mentionnés, la résolution des images est relativement faible.

En amont de la campagne d'acquisition de l'été 2013, nous nous sommes procurés des spécimens d'abeilles marquées et équipées d'une puce RFID. Cela nous a permis d'étudier la faisabilité de la détection vis-à-vis de la configuration de notre dispositif stéréovision (cf. Annexe D). Ces premiers tests ont démontré la faisabilité de la détection, tout du moins sur une majorité d'abeilles.

La figure 4.14 montre 2 exemples de traces issues d'une capture effectuée durant l'été 2013, l'une correspondant à une abeille pucée, l'autre correspondant à une abeille non pucée. Notre méthode de détection se décompose en 4 étapes. Sur chaque imagerie extraite à partir de la trajectoire :

1. localisation de la tête de l'abeille, correspondant à la zone la plus foncée de l'image et entourée par un gradient (méthode TFAC [Goldstein 1973] empruntée à la communauté RADAR) ;
2. localisation d'une zone claire à proximité de la tête, ce qui correspond soit à une puce RFID (si elle existe), ou soit à un élément aléatoire du fond de la scène ;
3. calcul de l'angle α_i et de la distance d_i entre la tête et la zone claire identifiée sur chaque imagerie i .

Puis, sur l'ensemble de la trajectoire :

4. calcul de deux indicateurs : la moyenne des différences absolues d'angles et de distances entre les couples d'imagettes successives, soit $\sum_i^{N-1} |\alpha_i - \alpha_{i+1}| / (N - 1)$ et $\sum_i^{N-1} |d_i - d_{i+1}| / (N - 1)$.

Ainsi, si ces indicateurs sont en dessous d'un certain seuil, on juge que l'abeille est porteuse d'une puce, sinon, on juge qu'elle ne l'est pas. Une vérité terrain sur 20 000 trajectoires, comportant plus de 600 abeilles pucées, a été construite. Des testes préliminaires ont montré des résultats intéressants en termes de détection de puce. Cependant, la méthode souffre encore d'un grand nombre de fausses détections. Celles-ci peuvent être facilement balayées à l'aide d'une vérification manuelle.

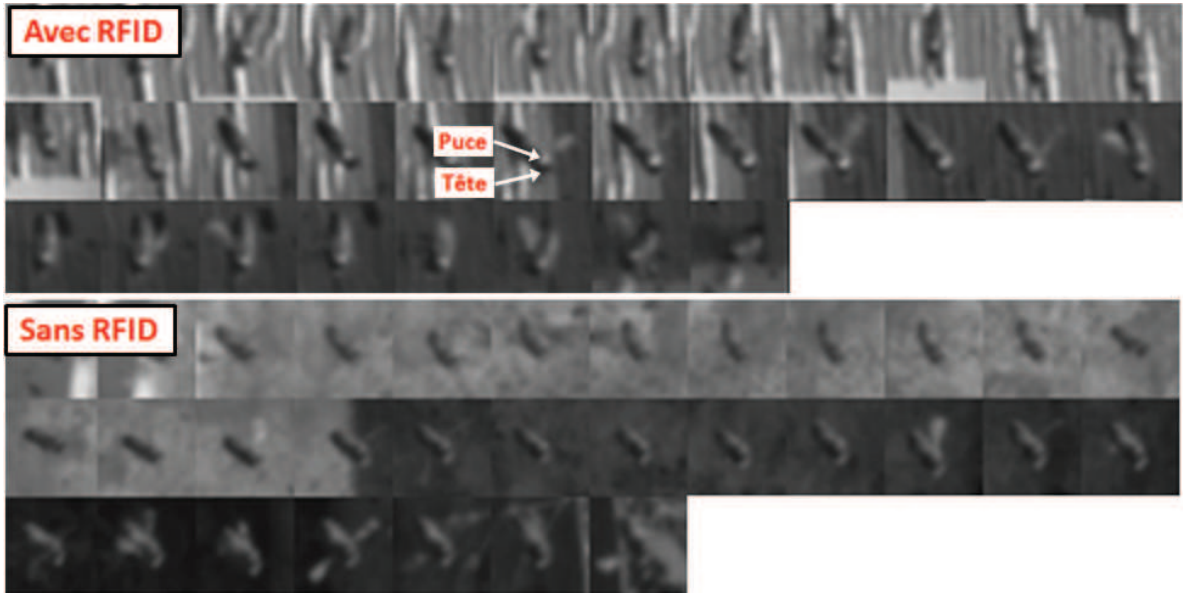


Figure 4.14 – Exemples d'imagettes extraites à partir de 2 trajectoires d'abeilles avec et sans puce RFID.

4.5.3 Nettoyage des données RFID

Lors de leurs entrées/sorties au niveau de la ruche, les abeilles sont canalisées successivement sous 2 rangées de lecteurs RFID permettant ainsi leur identification et la détection de leur sens de passage. Nommons "A" la rangée de lecteurs internes, et "B" la rangée de lecteurs externes. Une abeille passant successivement sous un lecteur de la rangée "A" puis sous un lecteur de la rangée "B" produit la séquence RFID "AB", ce qui correspond donc à une sortie. Inversement, la séquence "BA" correspond à une entrée. Ainsi, en détectant les motifs "AB" et "BA" engendrés par les abeilles, on peut retracer l'histoire de vie de l'individu. À partir de l'histoire de vie, on peut par exemple déduire la fréquence et le temps de sortie en fonction de l'âge. La figure 4.15 illustre le dispositif RFID.

Ce dispositif est loin d'être infallible. Les séquences RFID recueillies ressemblent difficilement à une succession de motifs "AB-BA" correspondant à un enchainement théorique d'entrées/sorties.

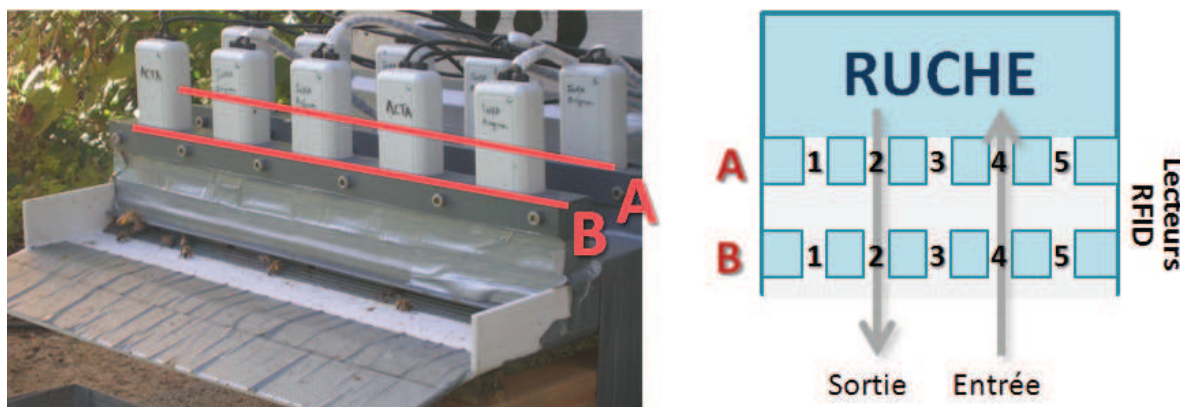


Figure 4.15 – À gauche : photo du système RFID constitué de 2 rangées de 5 lecteurs. À droite : schéma du système illustrant la disposition et la numérotation des lecteurs. Les abeilles possèdent une grande liberté de mouvement entre les 2 rangées de lecteurs.

- Il est assez fréquent qu'un lecteur ne détecte pas le passage d'une abeille. Le tag RFID peut être encrassé, défectueux ou mal orienté. Les abeilles peuvent parfois passer sens dessus dessous sous un lecteur ;
- Lorsqu'une abeille avance trop lentement, le lecteur RFID capte plusieurs fois la puce pour un même passage ;
- Il peut arriver qu'une abeille fasse demi-tour sous un lecteur RFID pour aller ensuite passer sous un lecteur voisin.

Ainsi, dans l'état, ces données sont difficilement exploitables. Les biologistes qui utilisent ce type de dispositif se contentent d'extraire les motifs "ABBA", ce qui ne concerne au final que 40% des données. Autrement dit, 60% des données collectées ne sont pas exploitées par les biologistes à cause des dysfonctionnements des lecteurs RFID.

Les pertes de données issues des différentes étapes de traitement commencent donc dangereusement à s'accumuler. C'est pourquoi il est nécessaire de débruiter et de compléter les séquences RFID en essayant par exemple de retrouver les détections manquantes. Pour cela, nous avons développé une méthode s'appuyant sur un algorithme génétique [Banzhaf 1998]. L'idée est que chaque individu représente un ensemble de règles (ou filtres) successives à appliquer à la séquence RFID. Le but de l'approche est de trouver un ensemble de règles qui transforme la séquence RFID brute en une histoire de vie cohérente. L'estimation de la cohérence d'une séquence est donnée par une fonction de fitness s'appuyant sur la différence entre la moyenne et la médiane, sur des intervalles de temps, des valeurs de durées (e.g. temps passé en dehors de la ruche pour un âge donné). L'ensemble de règles générant les individus est généralement simple. Un exemple de règle peut être l'ajout d'un "B" sur la séquence "ABA" sous certaines conditions pour qu'elle corresponde à un motif "ABBA".

Nous ne rentrons pas plus en détails sur le nettoyage des données RFID car cela s'écarte du cadre initial de cette thèse. En revanche, nous invitons les lecteurs à se référer au rapport de Master de Heriniaina Raharimanana (cf. [Raharimanana 2014]) pour plus de précisions.

4.5.4 Modalités d'application pour EPERAS

À la date de rédaction de ce manuscrit, l'implémentation de la chaîne de fusion d'information est encore incomplète. Les différents modules présentés précédemment sont fonctionnels, mais l'étape de synchronisation des horloges (cf. Section 4.3.4) est encore en cours de développement. Les sections suivantes présentent plusieurs approches méthodologiques qui seraient intéressant de mettre en œuvre sur les données issues de futures campagnes d'acquisition.

4.5.4.1 Trajectoires d'abeilles pucées VS non-pucées

Une première étude consisterait en la comparaison des trajectoires pucées et non-pucées au sein de l'espace des signatures. Par exemple, l'étude [Streit 2003] datant de 2003 mentionne que le vol des abeilles n'est pas affecté par le port de la puce RFID. La méthodologie ainsi que les outils développés permettraient ainsi de vérifier cela de manière directe sur des statistiques de plus grande ampleur.

La figure 4.16 illustre sur une capture, parmi l'ensemble des abeilles au départ de la ruche, quelques trajectoires d'abeilles pucées détectées automatiquement via la méthode proposés dans la section 4.5.2. Pour le moment, les numéros d'identification mentionnés sur le graphique sont simplement des identifiants de trajectoires 3D. Ils devraient à terme, après mise en correspondance des données, permettre de retrouver l'identité des abeilles pucées et ainsi relier les trajectoires en question aux différents niveaux d'intoxication correspondant.

Cela offre également des possibilités intéressantes en termes de comparaison. Par exemple, il serait alors possible d'étudier l'évolution des trajectoires d'une abeille tout au long de sa vie, ou bien, de comparer les trajectoires entre abeilles relativement à des paramètres donnés (e.g. âge, contamination).

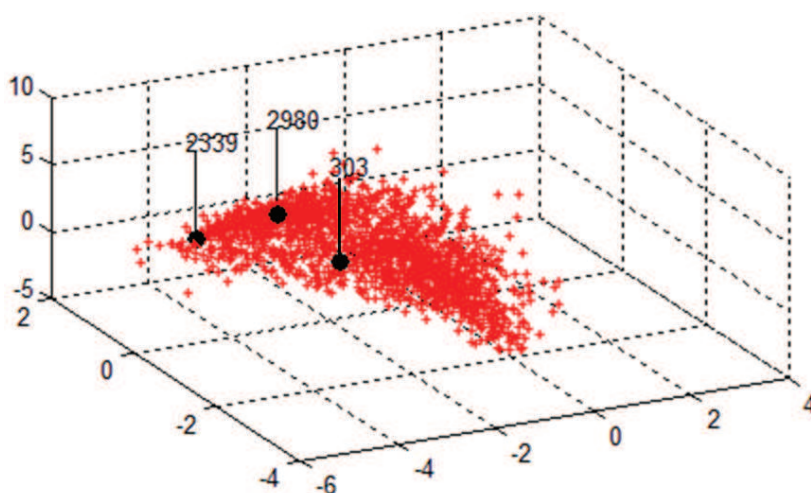


Figure 4.16 – Ensemble des signatures de trajectoires correspondant à des départs issues de la séquence capturée le 09/07/13 à 17h30. Chaque trajectoire est représentée par un point dans l'espace 3D issu d'une ACP sur l'espace des caractéristiques originales. Les 3 points noirs mettent en avant les 3 traces correspondant à des abeilles pucées. Celles-ci ont été identifiées automatiquement via la méthode de détection de puce (cf. Section 4.5.2).

4.5.4.2 Structuration du jeu de données selon l'intoxication

Une seconde étude directement rendue possible par la fusion d'information est une modélisation comportementale s'appuyant sur la méthode MLC-HDP (cf. Chapitre 3). Mais contrairement à la structure de données organisée en "*capture* (niv. 3) et *sousclips* (niv. 2)" utilisée dans la section 4.4.4, nous proposons de structurer le jeu de données en "*cohortes/niveaux d'intoxication* (niv. 3) et *sousclips* (niv. 2)". Si des changements de comportement s'avèrent réels, on peut alors espérer au travers de ce modèle, voir apparaître des comportements similaires pour toutes cohortes ayant subies les mêmes conditions d'intoxication.

4.6 Travaux annexes sur la détection du frelon asiatique

Cette section aborde brièvement deux études effectuées au L3i menées de manière complémentaire par rapport aux méthodes proposées dans cette thèse. L'une porte sur l'identification du frelon asiatique à partir des trajectoires collectées devant la ruche. L'autre porte sur l'analyse de l'activité des abeilles sur la planche d'envol.

4.6.1 Détection du frelon par sa taille

La première étude, menée conjointement avec Olivier Chappe, s'inspire de la méthode de caractérisation des cibles détectées s'appuyant sur une fonction d'appartenance à la classe "abeille" (cf. Section 1.4.3). Mais contrairement à la méthode mentionnée, l'objectif ici n'est pas de définir le degré de vérité d'une observation (i.e. chances d'être une fausse alarme), mais de chercher à savoir si la cible observée est une abeille ou un frelon asiatique. Pour cela, nous nous appuyons sur le rapport taille/profondeur de l'ellipse englobant la cible détectée. Comme le montre la figure 4.17, nous pouvons clairement distinguer deux modèles, l'un correspondant aux frelons, et l'autre correspondant aux abeilles. Nous utilisons une méthode de classification pour retrouver, à partir de l'ensemble des observations associées à une trajectoire, à quelle classe la trajectoire en question appartient.

L'objectif initial de ces travaux est l'estimation du taux de prélèvement d'abeilles par le frelon asiatique à l'entrée de la ruche. Le second intérêt de ces travaux est la construction d'une vérité terrain de la pression exercée sur la colonie par le frelon asiatique en fonction du temps. Celle-ci pourrait être utilisée pour évaluer notre méthode d'analyse afin d'étudier les modifications comportementales engendrées par cette pression.

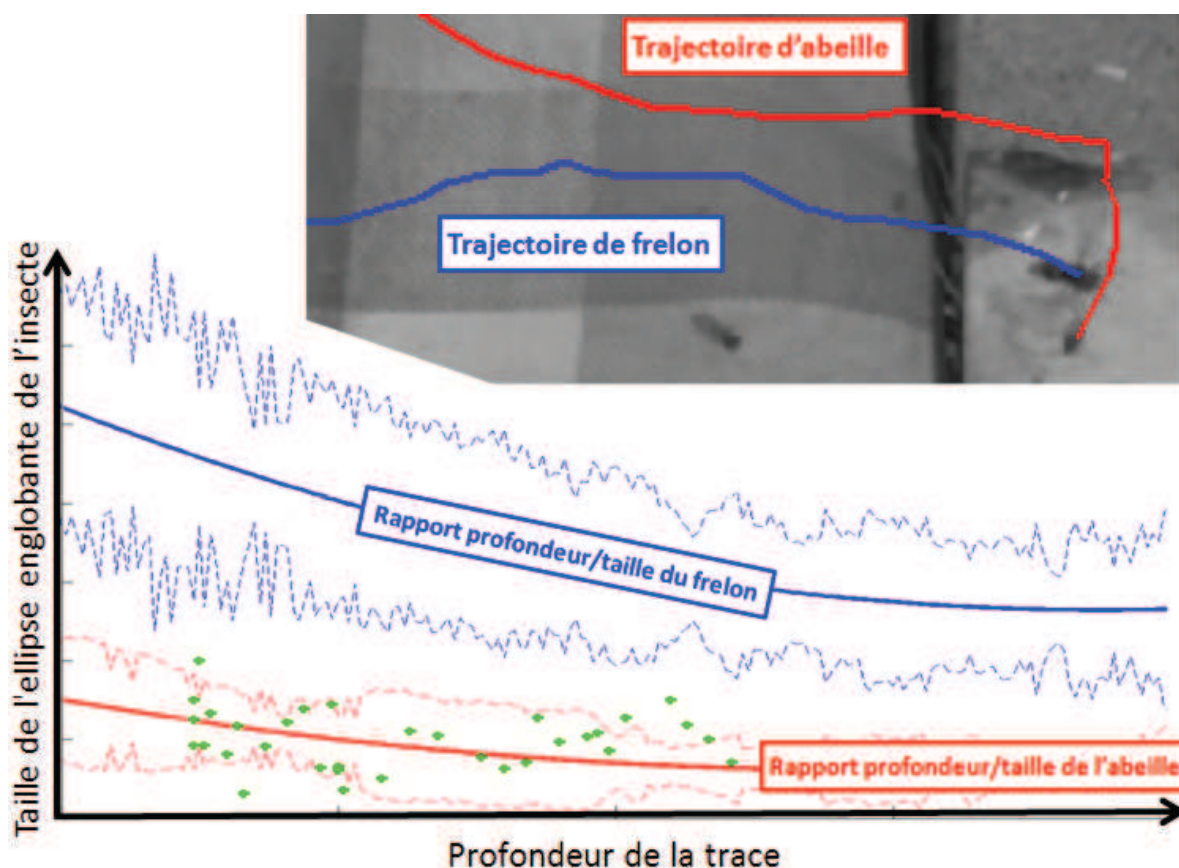


Figure 4.17 – En haut, deux trajectoires (une abeille en rouge et un frelon en bleu) classées automatiquement à partir des modèles. En bas, les deux aires modélisent de rapport profondeur/taille des ellipses englobantes, l'une correspond à la classe "abeille" (en rouge), l'autre correspond à la classe "frelon" (en bleu). Les points verts présents autour de la courbe rouge illustrent l'ensemble des observations liées à la trajectoire d'abeille montrée dans la partie supérieure de la figure. Les courbes pleines représentent les rapports moyens, et les courbes en pointillés représentent les écarts-types des modèles.

4.6.2 Analyse de l'activité sur la planche d'envol

La deuxième étude, menée par Clément Dorffer [Dorffer 2014], s'intéresse à l'activité des abeilles sur la planche d'envol. Les regroupements d'abeilles sur la planche d'envol donnent lieu à des scènes denses pour lesquelles les approches présentées dans cette thèse ne sont d'aucune utilité. Ainsi, dans cette étude, nous avons cherché à caractériser le mouvement des abeilles à l'aide d'une méthode de flot optique (cf. Figure 4.18). L'objectif est d'identifier des motifs ou des ruptures comportementales (au sens mouvement) sur la planche d'envol relativement au niveau de pression exercée par le frelon asiatique sur la colonie.



Figure 4.18 – Étude du mouvement sur la planche d'envol relativement au niveau de pression exercé par le frelon asiatique sur la colonie.

4.7 Conclusion

Ce quatrième et dernier chapitre, et plus généralement la thèse dans son ensemble, s'inscrit finalement comme une étude de faisabilité. Nous nous sommes appuyés sur les deux projets EPERAS et RISQAPI (projets interdisciplinaires composés d'experts de l'abeille ouverts à des approches d'investigation innovantes), pour tester la mise en œuvre, sur le terrain et en conditions réelles, de notre chaîne d'acquisition et de traitement de données. Pour rappel, voici les objectifs des deux projets :

- **EPERAS** avec l'INRA Magneraud comme partenaire principal, a pour cadre une étude toxicologique portant sur l'influence de produits phytosanitaires, assimilés par les abeilles à des doses sublétales, sur leur comportement.
- **RISQAPI** avec le Muséum National d'Histoire naturelle dans le cadre d'une étude portant sur le changement de comportement des abeilles face à la prédation par le frelon asiatique.

À l'occasion des expériences menées sur le terrain en conditions réelles, nous avons identifié de nouveaux verrous liés notamment à l'utilisation conjointe des données de trajectoires et de l'historique de vie des abeilles. Nous avons donc proposé, en plus d'une simple application de notre chaîne d'acquisition et traitement de données (cf. chapitres 1,2 et 3), des briques supplémentaires permettant de replacer les données de trajectoires collectées dans un contexte

biologique (e.g. histoire de vie des abeilles grâce aux capteurs RFID, activité sur la planche d'envol). Certaines des méthodes proposées sont encore en cours de développement, néanmoins la méthodologie permettant de lever l'ensemble des verrous a pu être évaluée. C'est ainsi que, par exemple, la méthode de fusion des données de trajectoires 3D et des données RFID, capitale pour la réussite du projet EPERAS, sera adoptée grâce aux verrous levés lors de la définition du protocole expérimental de la campagne de collecte des données de 2014/2015.

Dans le cadre du projet RISQAPI, nous avons pu tester notre approche de modélisation comportementale sur des données réelles. Les résultats de modélisation de comportements via le MLC-HDP sont loin d'être aboutis d'un point de vue biologique. Cependant, le fait que l'on observe une certaine stabilité des comportements dans le temps montre la cohérence de la méthode.

Finalement, on s'aperçoit que la quantité de données collectée lors de la campagne d'acquisition de 2013, bien que conséquente d'un point de vue informatique par la mise au point d'algorithmes, semble s'avérer insuffisante pour émettre des conclusions biologiques en termes d'interprétations comportementales. Cela sort d'ailleurs du cadre de cette thèse. Des études complémentaires doivent être menées de manière conjointe, et ceci sur plusieurs campagnes d'expérimentation.

Conclusion générale et perspectives

Aspects "vision par ordinateur" et "traitement du signal"

Dans ce manuscrit, nous avons proposé une chaîne complète d'acquisition et de traitement de données de trajectoires (cf. Figure 3) pour in fine, modéliser le comportement de colonies d'abeilles. Cette chaîne est constituée des maillons suivants, dans l'ordre :

- acquisition stéréovision ;
- segmentation RGB-D pour la détection de cibles ;
- suivi 3D multi-cibles pour le calcul des trajectoires ;
- extraction des signatures à partir des trajectoires individuelles ;
- structuration hiérarchique des données de trajectoires en entités temporelles ;
- modélisation comportementale par clustering multi-niveaux.

La construction de cette chaîne et sa mise en œuvre ont fait naître des problématiques inhérentes à plusieurs domaines. Après étude de la littérature correspondant à chacun de ces domaines, nous avons identifié les verrous scientifiques et technologiques à lever (cf. Tableau 1). Ces derniers nous ont amené à proposer différentes contributions, dont certaines ont donné lieu à des publications dans des journaux ou des conférences internationales (cf. Publications). Ces contributions sont inventoriées ci-après dans les sections correspondant aux différents blocs de notre chaîne d'acquisition et de traitement de données.

Dans le cadre de cette thèse, nous avons choisi de suivre une approche de travail horizontale. Autrement dit, notre priorité fut la construction d'une chaîne complète, plutôt que l'étude approfondie d'un bloc en particulier. De cette manière, nous avons pu valider d'un point de vue méthodologique, la solution proposée dans son ensemble. Cela s'est traduit par le développement et le test sur le terrain d'un prototype fonctionnel. Dans le futur, il serait intéressant de s'attaquer à l'amélioration des différents blocs de la chaîne. Quelques idées ont déjà été abordées au travers des différents chapitres de ce manuscrit. Ces idées ainsi que d'autres sont reprises dans les parties "Perspectives" des sections suivantes, chacune correspondant à un maillon de la chaîne. Quant aux différentes parties "contributions", elles résument brièvement les apports de cette thèse vis-à-vis de la littérature.

Dispositif d'acquisition RGB-D

Contributions Nous avons montré la faisabilité et proposé une configuration viable d'un dispositif stéréovision pour la capture et l'appréhension 3D de scènes encombrées (cibles vives, nombreuses, de petite taille et d'apparence similaire) dans le cadre d'applications de vidéosurveillance courte portée.

Perspectives Il est évident que l'amélioration constante des périphériques d'acquisition RGB-D doit être prise en compte dans le cadre d'une future mise en application de notre

chaîne d'acquisition et de traitement. En 2012, lors de l'achat de la caméra stéréovision, nous avons choisi celle qui offrait le meilleur compromis "fréquence d'acquisition/résolution". Depuis, le constructeur TYZX s'est fait racheter par Intel Corporation, ainsi la caméra en question (modèle G3 Evo) n'est plus disponible à la vente. En revanche il existe aujourd'hui très probablement des dispositifs offrant des compromis équivalents voire meilleurs. Il serait intéressant d'étudier les possibilités qu'offrent des dispositifs infrarouges à courte portée tels que le "Leap Motion" (cf. Section 1.2.2) dans le cadre de notre application de suivi de cibles.

Détection de cibles en 3D

Contributions Nous avons proposé une méthode de détection de cibles à partir de séquences RGB-D, que nous avons nommée "Segmentation HIDS" (cf. [Chiron 2013a]). Cette dernière exploite une triple fusion d'information : 1) mouvement par différence d'images d'intensité successives ; 2) mouvement par soustraction d'arrière plan d'intensité ; 3) soustraction d'arrière plan de profondeur.

Perspectives Notre méthode de segmentation HIDS permet de détecter la présence de cibles en vol sur une image issue d'une séquence vidéo RGB-D. Compte tenu de la nature des données collectées et de l'encombrement trop important de la planche d'envol, nous avons pris la décision de ne pas chercher à détecter les cibles situées à proximité de l'arrière plan (e.g. les abeilles posées sur la planche d'envol). À cause de cela, nous avons été confrontés dans le chapitre 3 à un manque de données entre la fin (à l'atterrissage) — ou le début (au décollage) — d'une trajectoire et le passage de l'abeille sous les lecteurs RFID. Il serait donc intéressant de reconsidérer la question et de voir si, dans des conditions particulières, il serait possible d'assurer la détection des abeilles isolées en déplacement sur la planche d'envol.

Suivi 3D multi-cibles

Contributions Nous avons proposé une méthode de projection anticipée permettant au filtre de Kalman (défini en 3D) de supporter un mélange d'observations définies en 2D et en 3D (cf. [Chiron 2013a]). Également, nous avons proposé deux méthodes distinctes s'appuyant sur des approches fondamentalement opposées, pour la prise en compte de l'influence de la scène sur la dynamique des cibles.

- Une méthode avec un fort *a priori*, s'exécutant rapidement : approche s'appuyant sur une relation hypothétique entre la dynamique d'une cible et sa proximité avec un élément (ou zone d'intérêt) de la scène (cf. [Chiron 2013b]).
- Une méthode avec peu d'*a priori*, demandant plus de ressources de calcul : approche s'appuyant sur une base de trajectoires (ou historique), représentant l'ensemble des dynamiques possibles en tous points de la scène (cf. [Chiron 2014a]).

Perspectives La dernière approche n'utilise plus l'étape de prédiction du filtre de Kalman, ce qui est plutôt drastique. On pourrait envisager de rendre la méthode plus générique, en utilisant une fusion entre la prédiction traditionnelle et la prédiction effectuée exclusivement sur les données.

Extraction des signatures

Contributions Plutôt qu'une approche fondée sur le flot optique telle que traditionnellement utilisée pour l'analyse de scènes encombrées, nous avons proposé d'appuyer notre modèle sur des trajectoires individuelles. Ceci a nécessité la détermination d'une signature de trajectoire adaptée à notre problème.

Perspectives Le calcul d'une signature optimale n'a pas été un point sur lequel nous avons souhaité nous attarder. En effet, nous avons proposé une signature minimale (i.e. invariants classiques sur la trajectoire). Bien que cette dernière ait permis d'obtenir des résultats significatifs concernant la modélisation comportementale, il serait intéressant d'étudier l'impact du choix des caractéristiques sur les résultats finaux.

Structuration hiérarchique des données

Contributions L'ensemble des signatures de trajectoires sont intégrées dans une structure hiérarchique organisée selon deux échelles temporelles différentes : (i.e. *captures* et *sous-clips*). A une large échelle, les entités de type "*capture*" contiennent des signatures issues de quinze minutes de vidéo, tandis que les entités de type "*sousclip*", à une échelle plus fine, contiennent des signatures issues d'une minute de vidéo.

Dans notre contexte applicatif, cette structuration a l'avantage de distinguer les événements ponctuels (e.g. attaque de frelon, essaimage) que l'on a désigné par le terme "*activité*", par rapport à des événements pouvant avoir des effets à moyen terme sur l'état de la ruche (e.g. mauvaise météo, maladies chroniques) que l'on a désigné par le terme "*comportement*".

Perspectives Tel que mentionné dans le chapitre 3, il serait opportun, dans le cadre de l'étude de toxicité du projet EPERAS de structurer les données différemment, par exemple en cohortes et en niveaux d'intoxication.

Modélisation comportementale par clustering multi-niveaux

Contributions À notre connaissance, les approches existantes de type HDP pour la modélisation comportementale utilisent au maximum 2 niveaux. Nous avons proposé d'utiliser le MLC-HDP, un HDP sur 3 niveaux de clustering, pour la découverte de comportements émergents à différentes échelles temporelles :

- *captures* similaires regroupées en *comportements* (à l'aide d'un DP) ;
- *sousclips* similaires regroupés en *activités* (à l'aide de plusieurs DPs, un par capture) ;
- trajectoires similaires regroupées *atomes d'activités* (à de plusieurs DPMs, un par *sous-clip*).

Pour la validation de notre méthode de modélisation, nous avons proposé d'utiliser une approche originale s'appuyant sur un SMA. L'utilisation d'un système multi-agents permet de fournir indirectement la vérité terrain sémantique associée au jeu de données de trajectoires. Cette approche originale de validation respecte les contraintes initiales de modélisation visant à introduire un minimum d'*a priori* dans le système.

Perspectives La méthode de clustering MLC-HDP permet de regrouper des données en injectant un minimum d'*a priori*. Il serait intéressant de prendre en compte l'aspect évolutif

des *comportements* dans le temps en étudiant par exemple les probabilités de transition entre états. De ce contexte, la méthode HDP-HMM (mentionnée dans la section 3.2.3.3) serait une solution à explorer.

Aspects applicatifs et biologiques

Les biologistes utilisent des outils de collecte et d'analyse de données toujours plus complexes. Nos travaux ont montré, au travers des différentes collaborations, que la vision par ordinateur a plus que jamais son rôle à jouer auprès des biologistes, que ce soit pour l'automatisation de tâches laborieuses ou bien pour la réalisation de tâches jusqu'à aujourd'hui inaccessibles.

D'un point de vue biologique, cette thèse s'avère être une étude méthodologique pour le développement d'outils visant à utiliser les trajectoires individuelles de vol des abeilles en tant que données comportementales exploitables. Non seulement nous avons proposé les méthodes permettant de lever la majorité des verrous scientifiques identifiés, mais nous avons également développé un prototype fonctionnel qui doit cependant être validé à plus grande échelle. Les projets EPERAS et RISQAPI vont encore se poursuivre quelques temps (fins prévues respectivement pour 2015 et 2016). Ainsi ils pourront bénéficier, dans le cadre de futures campagnes de collecte de données, des outils et des méthodes développées dans le cadre de cette thèse. D'ors et déjà, les forces et les faiblesses des outils proposés peuvent être prises en compte dans le design des protocoles expérimentaux.

Projet EPERAS

En ce qui concerne l'étude de toxicité du projet EPERAS, il reste à ce jour encore à implémenter le bloc permettant la synchronisation des horloges. Cela dit, les premiers résultats issus des 20 heures de vidéos capturées ont montré que le nombre d'abeilles pucées détectées par traitement d'images est relativement faible. Même dans le cas optimal (hautement improbable) où toutes les abeilles pucées détectées par RFID étaient mises en relation avec les trajectoires correspondantes, on estime que la quantité de données serait trop faible pour apporter une conclusion biologique. En biologie, l'étude du vivant et ses nombreuses variables associées nécessite un grand volume de données.

Perspectives Ainsi, dans le cadre d'une prochaine mise en œuvre du dispositif complet de vidéosurveillance couplé à des acquisitions RFID, il faudrait prévoir de pucer plus d'abeilles, et d'assurer des conditions optimales à tous les niveaux de la chaîne de traitement afin de perdre un minimum de données. En termes d'implémentation, il serait intéressant d'essayer de traiter ces données massives sur une plateforme de type Hadoop.

Projet RISQAPI

Concernant l'étude sur le frelon asiatique du projet RISQAPI, nous avons pu tester notre approche de modélisation comportementale sur des données où les attaques étaient simulées. Comme vu dans le chapitre 3, les ruptures comportementales potentiellement engendrées par nos perturbations artificielles de la colonie ne sont pas flagrantes. Même si l'expérience entreprise est finalement assez peu pertinente d'un point de vue biologique, on observe une

certaine stabilité des comportements identifiés dans le temps. Cela a contribué à démontrer la cohérence de la méthodologie dans son ensemble.

Perspectives La saison pendant laquelle on peut observer le frelon dans de bonnes conditions est assez restreinte. Nous n'avons pas eu l'opportunité de mettre en œuvre des manipulations plus élaborées. Nous envisageons de renouveler l'expérience avec un frelon captif que l'on approcherait d'une ruche pour étudier les perturbations au niveau de la colonie.

Pour conclure ce manuscrit, nous invitons les lecteurs qui souhaitent approfondir les aspects biologiques à consulter le manuscrit de thèse de Fabrice REQUIER¹¹. Fabrice nous a accompagné sur le terrain lors des nombreuses manipulations et il nous a conseillé sur le côté applicatif de cette thèse.

11. Docteur en Ecologie de l'abeille affilié à l'INRA.

Outils d'annotation

A.1 Annotation de cibles

L'outil d'annotation des cibles a été réalisé sous la forme d'une application web (html, javascript, php et mysql). Les participants s'identifient sur leur session utilisateur, et annotent l'ensemble des abeilles en vol à partir des informations 1) d'intensité, 2) de profondeur et 3) de mouvement. L'information de mouvement est donnée par l'affichage du couple d'images intensité/profondeur suivant au sein de la séquence. Ce dernier est directement temporairement superposé sur l'ancien couple, ce qui permet à l'utilisateur d'apprécier le mouvement. Ces 3 sources d'information sont indispensables à l'utilisateur pour pouvoir distinguer les abeilles en vol parmi les autres éléments présents dans la scène (e.g. ombres, abeilles sur la planche). La figure A.1 montre l'interface de l'outil.

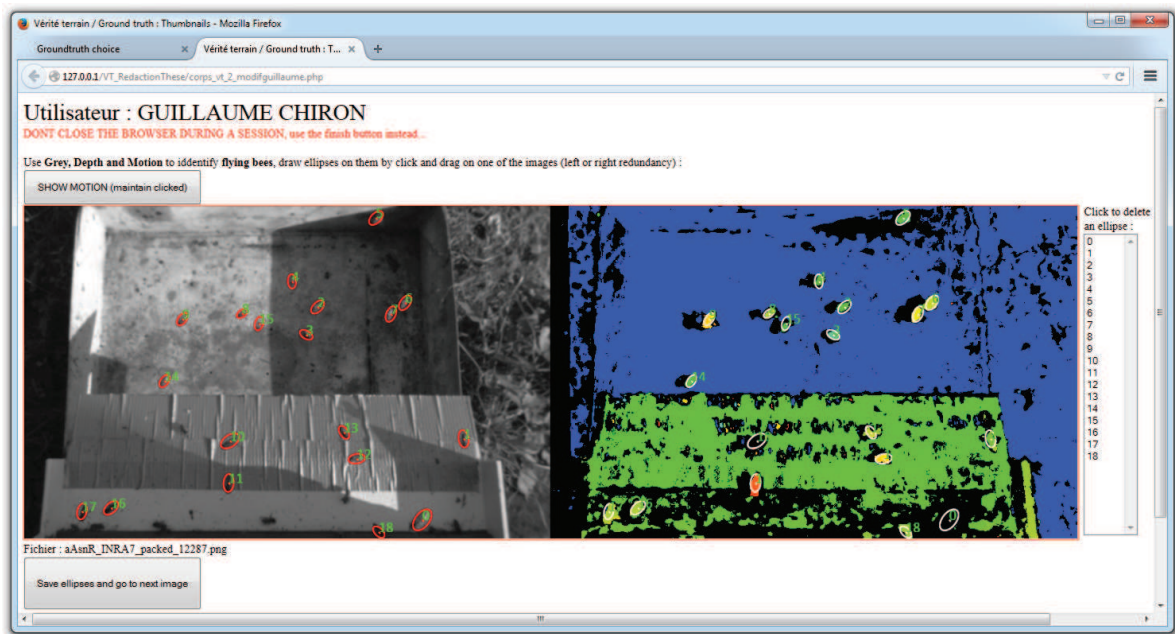


Figure A.1 – Capture d'écran de l'outil d'annotation de cibles. Image de gauche : information d'intensité ; image de droite : information de profondeur correspondante ; bouton "SHOW MOTION" : affiche le couple intensité/profondeur suivant au sein de la séquence ; ellipses : abeilles en vol annotées.

A.2 Annotation de trajectoires

L'outil d'annotation de trajectoires a été réalisé en Matlab. L'objectif de cet outil est de construire une base de données de trajectoires réelles. Lors de la construction de la vérité terrain pour l'évaluation de notre méthode de détection de cibles (cf. Section A.1), nous avons dû annoter l'ensemble des abeilles présentes dans chaque image issue d'un échantillon d'images tirées au hasard. Ici, nous nous contentons d'annoter l'ensemble des observations relatives à chaque trajectoire d'abeille d'un échantillon de trajectoires tirées au hasard. La figure A.2 montre l'interface de l'outil.

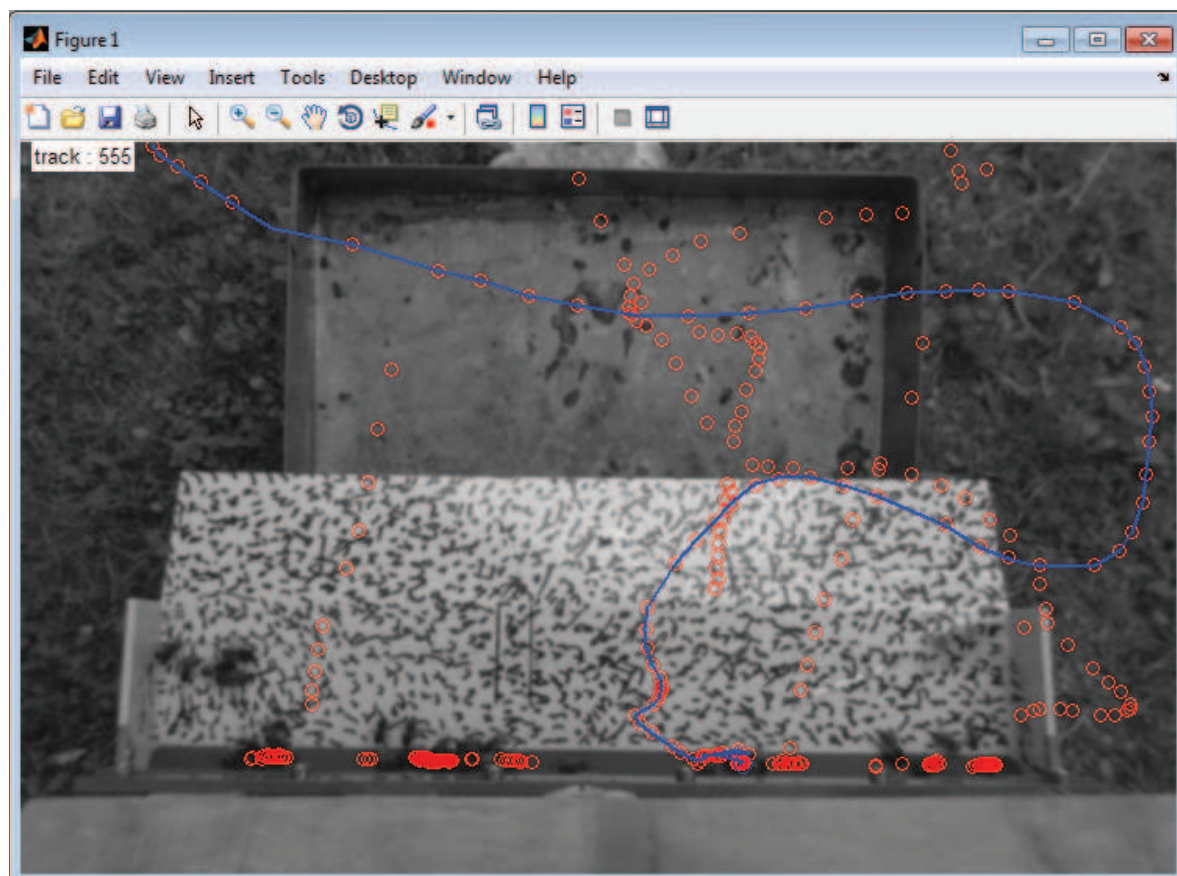


Figure A.2 – Capture d'écran de l'outil d'annotation de trajectoires. Pour aider l'utilisateur, l'outil affiche, sous forme de cercles, l'ensemble des cibles détectées à l'aide de la méthode de segmentation HIDS (cf. Section 1.4) sur un laps de temps assez large pour englober l'ensemble de la trace à annoter. Au fur et à mesure que la séquence se déroule, l'utilisateur clique successivement sur les cercles correspondant à la cible pour reconstruire la trace.

A.3 Détournage de la planche d'envol

L'outil d'annotation de la structure de la planche d'envol a été réalisé en Matlab. Cet outil permet à l'utilisateur de renseigner les coins de la planche d'envol, ce qui est nécessaire pour la reconstruction du maillage 3D de la planche (cf. Section 2.4.3). Ce modèle 3D est utilisé au sein des diverses étapes de la chaîne de traînement et d'analyse proposée.

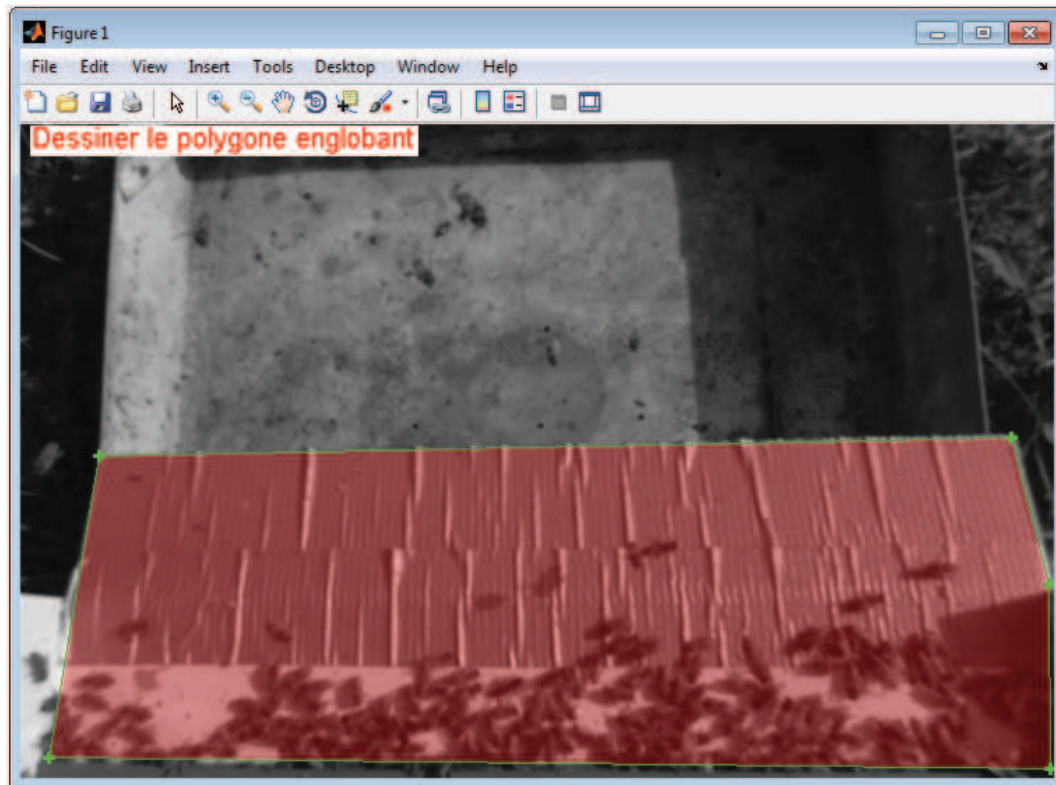


Figure A.3 – Capture d'écran de l'outil de détournage de la planche d'envol. L'utilisateur clique successivement (dans le sens horaire) sur les coins de la planche d'envol. Le polygone calculé correspondant à la zone s'affiche en transparence.

SMA de "Colonie d'abeilles"

Chaque agent "abeille" navigue entre les 3 statuts suivants :

- **"Exploratrice"** : explore le monde à la recherche d'une source de nourriture. Si l'exploration demeure infructueuse au delà d'une certaine période, l'abeille est contrainte de revenir à la ruche. Lorsque qu'une source de nourriture est trouvée, l'abeille peut devenir "Exploitante";
- **"Exploitante"** : exploite une source de nourriture en faisant des allers-retours entre la source et la ruche;
- **"Suiveuse"** : positionnée dans la ruche, elle accumule les informations sur les sources de nourritures partagées par les "Exploitantes" lors des multiples va-et-vient. Au bout d'une certaine période, la "suiveuse" peut devenir "Exploitante" et exploiter à son tour la source de nourriture la plus partagée par ses congénères.

Les principaux paramètres du modèle sont :

- Paramètres internes au modèle
 - **probaChangeStatus** définit les chances, pour l'abeille qui arrive à la ruche, de changer de statut;
 - **probaDépart** définit, à chaque itération, les chances que l'abeille quitte la ruche;
 - **distParcourueMax** définit la distance maximum qu'une exploratrice peut accomplir avant de devoir rentrer à la ruche.
- Paramètres de simulation d'intoxication
 - **probaPartage** définit si oui ou non la suiveuse va enregistrer l'information transmise par l'exploitante;
 - **probaDetectionNourriture** définit si oui ou non l'exploratrice va percevoir la source de nourriture.

(nommés respectivement "*Taux de Communication*", *TC*, et "*Taux de Sensibilité*", *TS* dans le manuscrit).


```

foreach Abeille do
    /* ----- ACTIONS À L'ARRIVÉE ----- */
    if arriveALaRuche then
        distParcourue = 0, porteNourriture = 0, resteAlaRuche = 1
        if "Exploitante" then
            forall the "Suiveuse" do
                if probaPartage then augmenteScoreDeLaSourcePartagée;
            end
        end
        if probaChangeStatus then
            statut = "Suiveuse"
            réinitialiseScoresSources
        end
    end
    if proximitéNourriture then
        if sourceVide then statut = "Exploratrice";
        else if probaDetectionNourriture then
            statut = "Exploitante", porteNourriture = 1
            définirLaSourceCommeMeilleurScore
        else
            statut = "Exploratrice"
            réinitialiseScoresSources
        end
    end
    /* ----- ACTIONS AU DEPART ----- */
    if probaDépart then
        resteAlaRuche = 0
        if statut = "Suiveuse" then
            if sourcesConnues then statut = "Exploitante"
            else statut = "Exploratrice";
        end
    end
    /* ----- DEPLACEMENTS ----- */
    if resteAlaRuche = 0 then
        if probaDetectionNourriture then
            if distParcourue < distParcourueMax then deplaceAleatoirement
            else deplaceVersRuche;
        else
            if statut = "Exploitante" then
                if porteNourriture = 1 then deplaceVersRuche
                else deplaceVersLaSourceAyantLeMeilleureScore;
            end
        end
    end
end
end

```

Algorithme 3 : Algorithme SMA "Colonie d'abeilles"

Fiche d'annotations

La figure C.1 présente un exemple de fiche de suivi des acquisitions effectuées par nos soins. Cette fiche a pour objectif de répertorier des informations supplémentaires (e.g. météo, activité, événements marquants) permettant par la suite, à l'image d'une vérité terrain, de retrouver certaines tendances potentiellement observables sur les résultats issus de l'analyse comportementale à partir des trajectoires.

Journée de manipulation

Répertoire des captures de la journée (modèle: /manip_jj-mm-yy) : /manip_12-12-13

Resp: Bruno

Heure (corriger si besoin)	Durée (corriger si besoin)	Sous Répertoire de capture	Activité (Forte, Moyenne, Faible)	Météo (Soleil, Couvert, Vent ...)	Autres observations (Attaque frelon (noter le temps si possible), problème caméra...)
14h00	15 min	/INRA1C1	Forte	Couvert + pluie légère	Beaucoup de mouches surtout derrière la roche.
14h30	15 min	/INRA2C2	Forte ⊕	Couvert	
15h00	15 min	/INRA3C3	//	//	//
15h30	15 min	/INRA4C4	// ⊕ ⊕	Couvert + éclaircies	//
16h00	15 min	/INRA5C5	//	Couvert + éclaircies ⊕	//
16h30	15 min	/INRA6C6	//	// // ⊕ ⊕	//
17h00	15 min	/INRA7C7	//	// //	//
17h30	15 min	/INRA8C8	//	Soleil Voilé	//

RAPPELS :

- 1) **CHANGER DE REPERTOIRE** avant d'enregistrer une nouvelle capture !!!
- 2) S'assurer que l'on capture en **MODE FULL** !!!

Dpc - RFP = 14h
Bac d'eau humide

Figure C.1 – Exemple de fiche de suivi des acquisitions vidéo journalières.

Détection visuelle des marqueurs

La petite taille du tag RFID rend sa détection particulièrement difficile. La figure D.1 illustre l'ordre de grandeur des spécimens étudiés et des marquages envisagés.

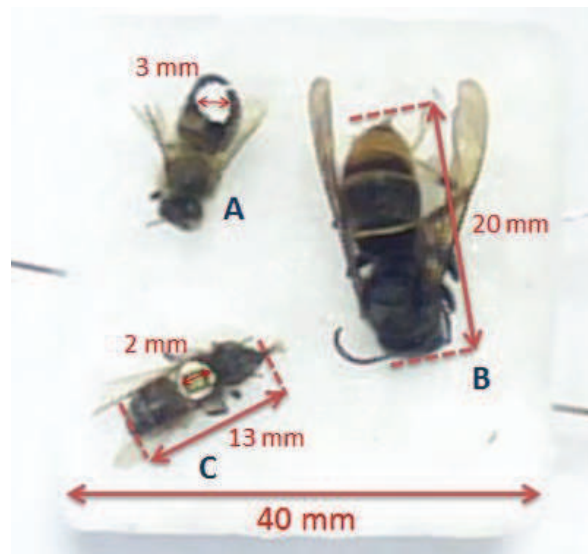


Figure D.1 – Ordre de grandeur des spécimens utilisés pour l'étude. A) Abeille marquée d'une tâche blanche (peinture marquante) d'une taille de 3 mm, B) Frelon asiatique mesurant 20 mm, C) Abeille (13 mm) équipée d'une puce RFID (2 mm), avec le plâtre blanc débordant sur les bords de la puce.

La figure D.2 montre pour les spécimens, différents triplets "image d'intensité / image RGB / carte de profondeurs" acquis par la caméra stéréovision TYZX G3 Evo placée à des distances bornées entre 42.5 cm et 60 cm.

La marque blanche offre des possibilités intéressantes en termes de détection et ce directement sur les images en niveaux de gris. Par ailleurs, la texture uniforme de la peinture ne perturbe pas le calcul de la carte de profondeur à cet endroit. À 42.5 cm, la marque mesure environ 9 px², et à 60 cm elle mesure 4 px². Il est préférable de rester en dessous des 52.5 cm de hauteur afin d'avoir assez de pixels pour détecter la marque. La détection du tag RFID sans marquage peinture supplémentaire peut être également envisagée à condition de laisser généreusement déborder le plâtre servant à la fixation de la puce.

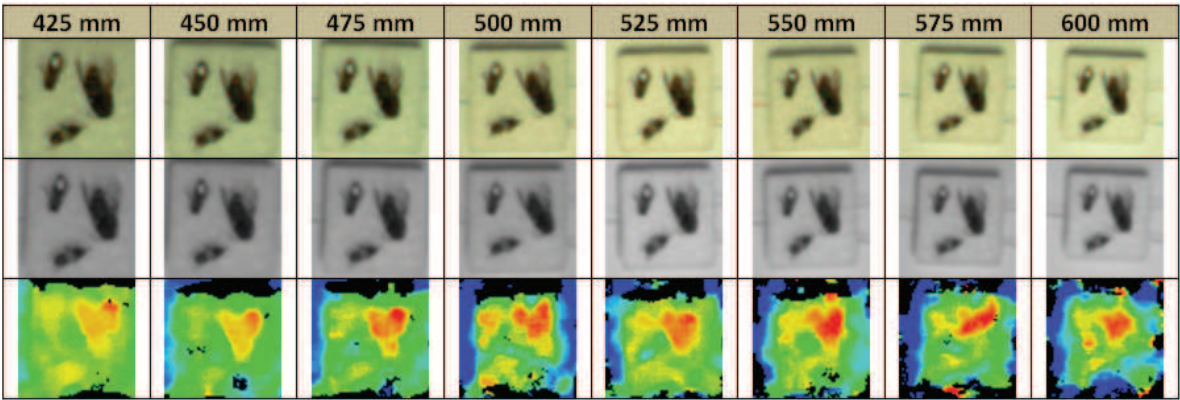


Figure D.2 – Spécimens vus par la caméra stéréo-vision TYZX G3 Evo placée à des distances bornées entre 42.5 cm et 60 cm. Première ligne : couleur, deuxième ligne : niveaux de gris, troisième ligne : carte de profondeur.

Publications

Articles de journaux

Internationaux

- Chiron G., Gomez-Krämer P., Michel M., Detecting and tracking honeybees in 3D at the beehive entrance using stereo vision, EURASIP Journal on Image and Video Processing, vol. 2013, page 59, 2013

Nationaux

- Chiron G., Gomez-Krämer P., Michel M., Modélisation comportementale selon différentes échelles temporelles à partir de trajectoires issues de scènes encombrées, Revue d'Intelligence Artificielle (RIA), 2014/2015 (*papier soumis le 31/10/2014*)

Articles de conférences & workshops

Internationaux

- Chiron G., Gomez-Krämer P., Ménard M., Discovering Emergent Behaviors from Tracks Using Hierarchical Non-parametric Bayesian Methods, 22nd International Conference on Pattern Recognition (ICPR), pages 2185-2190, Stockholm, 27 août 2014
- Chiron G., Gomez-Krämer P., Ménard M., 3D tracking based on possibilities rather than probabilities. Application to flying honeybees at the beehive entrance, In Workshop Visual observation and analysis of Vertebrate And Insect Behavior (VAIB), in conjunction with ICPR, Stockholm, 24 août 2014
- Chiron G., Gomez-Krämer P., Ménard M., Requier F., 3D Tracking of Honeybees Enhanced by Environmental Context, Dans Image Analysis and Processing, International Conference on Image Analysis and Processing (ICIAP), vol. 8156, pages 702-711, Naples, 13 septembre 2013
- Chiron G., Gomez-Krämer P., Ménard M., Outdoor 3D Acquisition System for Small and Fast Targets. Application to honeybee monitoring at the beehive entrance, In Workshop Proceedings of GEODIFF, in conjunction with VISIGRAPP, pages 10-19, Barcelone, 23 février 2013

Nationaux

- Chiron G., Gomez-Krämer P., Ménard M., Approche Bayésienne non paramétrique pour la découverte de comportements émergents à partir de trajectoires, 19ème congrès national sur la Reconnaissance de Formes et l'Intelligence Artificielle (RFIA), Rouen, 3 juillet 2014

Autres rassemblements scientifiques sans actes

Nationaux

- Chiron G., Gomez-Krämer P., Ménard M., Processus de Dirichlet hiérarchique pour la découverte de comportements. Validation sur des trajectoires d'insectes simulées par SMA, 7ème Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes (JFRB), Paris, 26 juin 2014, Article de conférence
- Chiron G., Gomez-Krämer P., Ménard M., Détection et suivi d'abeilles en 3D, Analyse conjointe des images RGB-Depth, GdR 720 ISIS, Paris, 6 février 2014, Présentation

Locaux

- Chiron G., Gomez-Krämer P., Ménard M., Suivi 3D multi-cibles & multi-hypothèses, Étude comportementale des abeilles, Colloque FREDD, Île d'Oléron, 28 juin 2012, Poster
- Chiron G., Gomez-Krämer P., Ménard M., Suivi 3D multi-cibles & multi-hypothèses, Découverte de comportements récurrents à différentes échelles temporelles à partir de trajectoires, Journée Reconnaissance d'actions et de comportements dans les vidéos, Fédération MIRES, La Rochelle, 3 octobre 2014, Présentation

Couverture médiatique

TV locale

- "Environnement : Le labo des abeilles" présentant conjointement la société Apilab¹ et les travaux sur les trajectoires menés au L3i, France 3 Atlantique, Journal télévisé édition local, 5 Septembre 2012, 19h20 (2 min. 09)

Articles de presse faisant référence à nos travaux

- "Des abeilles en 3D" paru dans "Vu bien vu", magazine du développement durable en Aquitaine, n°17, page 6, octobre 2012
- "NTIC Apialert" paru dans "Techni cités", le magazine des cadres techniques de la fonction publique, n°236, 8 octobre 2012

Animations locales et vulgarisation

- Tenue d'un stand lors de la Fête de la science, la Rochelle, éditions 2013 et 2014, [http ://www.fetedelascience.fr](http://www.fetedelascience.fr)
- Tenue d'un stand dans le cadre de la 8ème édition du concours national "Faites de la Science", la Rochelle, mais 2013, [http ://www.faitesdelascience.fr](http://www.faitesdelascience.fr)

1. A l'origine du système Apialerte, récompensé par : 1) la meilleure innovation aux World Beekeeping Awards 2013 (Kiev, Ukraine) ; 2) le lauréat 2013 du trophée de l'éco-entreprise innovante, dans la catégorie génie écologique (Bercy) ; 2) le prix de l'éco-innovation au salon Environord 2012 (Lille)

Liste des Abréviations

INRA	Institut National de la Recherche Agronomique
CEBC	Centre d'Études Biologique de Chizé
FREDD	Fédération de Recherche en Environnement pour le Développement Durable
EPERAS	Sigle du projet visant à développer un système utilisant l'abeille pour la surveillance environnementale de sites à risque technologique
RISQAPI	Sigle du projet visant à faire le lien spatial entre usage des pesticides et risques d'affaiblissement des colonies d'abeilles en système de grandes cultures
VAIB	Visual observation and Analysis of vertebrate and Insect Behavior
LIDAR	Light Detection And Ranging
RGB-D	Red, Green, Blue et Depth
TOF	Time Of Flight
HIDS	Hybrid Intensity Depth Segmentation
PF	Particle Filter
SIR	Simple Importance Resampling
KF	Kalman Filter
PHD	Probability Hypothesis Density
GNN	Global Nearest Neighbour
MHT	Multiple Hypothesis Tracking
DP	Dirichlet Process
DPM	Dirichlet Process Mixture model
HDP	Hierarchical Dirichlet Process
NDP	Nested Dirichlet Process
MLC-HDP	Hierarchical Dirichlet Process Model with Multiple Levels of Clustering
GEM	Griffiths Engen McCloskey
ACP	Analyse en Composantes Principales
CDF	Centred Distance Function
SMA	Système Multi-Agents
TD, TE	Taux de Diffusion, Taux d'Évaporation
TC, TS	Taux de Communication, Taux de Sensibilité
TFAC	Taux de Fausses Alarmes Constant
RFID	Radio-Frequency IDentification

Table des figures

1	Test du dispositif d'acquisition stéréovision dans le rucher du domaine du Magneraud durant l'automne 2012.	7
2	Illustration schématique présentant de manière extrêmement simplifiée le problème de la modélisation comportementale en fonction des facteurs en synergie susceptibles d'influencer le jeu de données de trajectoires.	8
3	Découpage en 4 chapitres de notre chaîne d'acquisition et de traitement de données.	12
1.1	Image d'intensité montrant une activité moyenne des vols à l'entrée d'une ruche.	18
1.2	Photos des deux caméras CamBoard Nano (TOF) et G3 Evo (Stéréovision) utilisées pour nos tests.	21
1.3	Comparaison des données acquises par 2 technologies RGB-D différentes. . . .	22
1.4	Illustration des phénomènes de "zones d'ombre" observés sur les cartes de disparité.	23
1.5	Configuration du dispositif d'acquisition stéréovision.	25
1.6	Différents effets observés sur la carte de profondeur issues de la caméra stéréovision TYZX G3 Evo.	29
1.7	Schéma illustrant la complémentarité intensité/profondeur pour la détection de cibles non appariées.	30
1.8	Illustration du calcul de DBG et $DDTM$ à partir d'un exemple de séquence capturée à l'entrée d'une ruche.	32
1.9	Illustration du calcul de $IAMM$ à partir d'un exemple de séquence capturée à l'entrée d'une ruche.	33
1.10	Illustration du calcul de $IRMM$ sur une séquence capturée à l'entrée d'une ruche.	34
1.11	Illustration du calcul de $UDTM$ et de FG sur une séquence capturée à l'entrée d'une ruche.	35
1.12	Algorithme de segmentation HIDS sous forme schématique.	36
1.13	Aperçu des étapes clés de l'algorithme HIDS sur une image issue d'une séquence RGB-D capturée devant la ruche.	36
1.14	Quatre exemples de cibles segmentées, avec les ellipses correspondantes sur les 4 couples d'images d'intensité/profondeur.	37
1.15	Relation moyenne longueur/profondeur pour une cible de type "Abeille".	38
1.16	Dispositif d'acquisition stéréovision fixé à l'entrée d'une ruche.	39
1.17	Images extraites des neuf séquences utilisées pour la construction de la vérité terrain.	40
1.18	Échantillons de 100 abeilles annotées à partir des neufs captures disponibles. .	41
1.19	Représentation de la distribution spatiale des erreurs de détection.	42

2.1	Caractérisation ambiguë du mouvement au niveau spatio-temporel.	46
2.2	Différentes approches de suivi de cibles.	48
2.3	Mécanisme récursif (Prédiction/Correction) du filtre de Kalman	50
2.4	Illustration du filtre SIR (Sampling Importance Resampling).	52
2.5	Illustration des différentes méthodes d'association de cibles, naïves (à gauche) et plus évoluées (à droite). Figure adaptée de [Yilmaz 2006].	56
2.6	Observation exprimée dans le repère image et dans le repère du monde réel. . .	60
2.7	Mécanisme de projection local en cas d'absence d'information de profondeur. .	61
2.8	Mécanisme de reprojexion suite à l'absence d'information de profondeur pendant un laps de temps donné.	61
2.9	Estimation du bruit l'état contraint par une zone d'intérêt de la scène.	63
2.10	Illustration des différentes étapes du processus de prédiction effectué sur la base de données de trajectoires.	67
2.11	Erreur sur l'état prédit relativement à la taille de la base de données de trajectoires.	68
2.12	Aperçu des problèmes à résoudre pour la mise en œuvre du suivi d'abeilles en 3D devant la ruche à partir des observations issues de notre segmentation HIDS. .	69
2.13	Aperçu, sur un échantillon de traces d'abeilles, des coûts d'associations par rapport à la distance à la planche d'envol.	70
2.14	Aperçu de la planche d'envol (vue du dessus, tournée d'un quart de tour) en cours de reconstruction.	72
2.15	Écart-types et moyennes des valeurs de profondeur de la planche d'envol durant le processus itératif de reconstruction.	72
2.16	Planche d'envol légèrement incurvée reconstruite sous forme de maillage 3D. . .	73
2.17	Illustration de 0.1% des trajectoires constituant la base de données.	73
2.18	Exemples de trajectoire contenues dans la vérité terrain. Les marqueurs plus épais présents sur les trajectoires indiquent les observations originalement dépourvues d'information de profondeur, pour lesquelles ces informations manquantes ont été interpolées.	74
2.19	Comparaison des temps de calcul entre les méthodes GNN et MHT.	77
2.20	Étude des performances 2D, 3D des approches GNN et MHT.	78
2.21	Étude des performances de l'approche 3D-GNN, avec et sans, utilisation de la méthode de re-projection.	79
2.22	Étude de la robustesse de l'approche 3D-GNN sous différentes conditions de segmentation. Les expériences ont été réalisées sur $3 * 9 * 100 = 2700$ scénarios distincts de type "Normal", "FA+" et "MD+".	80
2.23	Comparaison des 2 méthodes de suivi contraintes, l'une avec modèle de dynamique (Contrib. 2a) et l'autre s'appuyant principalement sur les données (Contrib. 2b).	81
2.24	Étude des performances apportées par nos Contrib. 2a et 2b (cf. Section 2.3) relativement à une approche classique. Les expériences ont été réalisées sur $6 * 100 = 600$ scénarios distincts de type "Couple".	81
2.25	Expériences pour le paramétrage empirique des modèles de bruit (q et r) du filtre de Kalman.	82

3.1	Exemple de fonctions de densité de probabilité de la loi de Binomiale et de sa conjuguée, la loi Beta.	89
3.2	Exemple de fonctions de densité de probabilité : loi multinomiale et sa conjuguée, la loi de Dirichlet.	90
3.3	Exemple où H est une loi de probabilité normale.	91
3.4	Construction stick-breaking du support de la distribution discrète G à partir de la distribution H	92
3.5	Représentation graphique et équations d'un DPM.	92
3.6	Comparaison de la méthode K-means et DP sur des données générées aléatoirement à partir d'un mélange de gaussiennes.	93
3.7	Représentation graphique et équations d'un HDP.	94
3.8	Illustration de la faculté de partage des clusters de la méthode HDP sur des données structurées en groupes, comparée à l'approche DP classique.	94
3.9	Représentation graphique et équations d'un MLC-HDP à deux niveaux.	95
3.10	Comparaison des 4 modèles : Processus de Dirichlet (DP), le processus hiérarchique de Dirichlet (HDP), le processus de Dirichlet imbriqués (NDP) et le clustering HDP multi-niveaux (MLC-HDP).	96
3.11	Structure de données hiérarchique organisée " <i>captures, sousclips, trajectoires</i> ".	98
3.12	Illustration du principe d'inférence des paramètres du modèle à partir des données.	100
3.13	Exemple de calcul de Fonction de Distance Centrée (CDF) à partir d'une trajectoire 2D.	101
3.14	Progression de la log vraisemblance négative au cours des itérations.	102
3.15	Schéma des trajectoires engendrées par un agent à l'issue d'une simulation SMA.	104
3.16	Simulations Netlogo de l'algorithme des fourmis selon 4 configurations de paramètres.	106
3.17	Signatures (projetées dans un repère 3D via une ACP) pour deux capture simulées selon des configurations différentes. Le modèle multi-agents utilisé est celui des fourmis présenté dans la section précédente.	107
3.18	Analyse des résultats de clustering MLC-HDP (niv. 2) vis-à-vis de la vérité terrain pour le modèle des fourmis.	108
3.19	Analyse des résultats de clustering MLC-HDP (niv. 3) vis-à-vis de la vérité terrain pour le modèle des fourmis.	109
3.20	Automate utilisé pour le modèle "Abeilles", modélisant les conditions de transition entre les 3 statuts que peuvent prendre les agents.	111
3.21	Vue 3D Netlogo d'une simulation en cours (à la 300ème itération) de notre modèle multi-agents de colonie d'abeilles.	111
3.22	Ensemble des trajectoires 3D contenues dans chaque <i>sousclips</i> (SC_n) d'une <i>capture</i>	113
3.23	Analyse des résultats de clustering MLC-HDP (niv. 2) vis-à-vis de la vérité terrain pour le modèle des abeilles.	114
3.24	Résultat de clustering (niv. 1) sur l'ensemble des signatures de trajectoires contenues dans une <i>capture</i>	115
3.25	Analyse des résultats de clustering MLC-HDP (niv. 3) vis-à-vis de la vérité terrain pour le modèle des abeilles.	117

4.1	Principe de partage des données de la plateforme Preserve.	124
4.2	Calendrier effectif des interventions 2013 conjointes L3i/INRA, principalement pour l'étude toxicologique liée à l'abeille domestique prévue dans le cadre du projet EPERAS.	127
4.3	Dispositif d'acquisition stéréovision à l'entrée de la ruche.	128
4.4	Données acquises selon trois horloges différentes.	129
4.5	Exemple de détection d'abeilles en vol via notre méthode de segmentation HIDS en condition de forte activité	131
4.6	Exemple de calcul de trajectoires d'abeilles en vol en condition de forte activité.	133
4.7	Projection dans un espace 3D des signatures de l'ensemble des trajectoires d'une capture.	134
4.8	Structuration des signatures de trajectoires au sein d'un jeu de données hiérarchisé.	136
4.9	Résultat du clustering MLC-HDP (niv. 1 - <i>atome d'activité</i>), sur l'ensemble des données de trajectoires, couvrant la totalité des <i>captures</i> acquises durant l'été 2013 à l'INRA Magneraud.	137
4.10	Résultat de clustering MLC-HDP (niv. 2 - " <i>activité</i> ") sur les 21 derniers <i>sous-clips</i> d'une journée.	138
4.11	Résultat de la modélisation MLC-HDP (niv. 3 - " <i>comportements</i> ") sur l'ensemble des 74 <i>captures</i> acquises durant l'été 2013 à l'INRA Magneraud dans le cadre de l'étude toxicologique.	139
4.12	Synthèse des performances (en termes de temps de calcul) constatés lors de l'application des différentes étapes de la chaîne.	140
4.13	Illustration du problème de raccordement des données visuelles de trajectoires et les données RFID	142
4.14	Exemples d'images extraites à partir de 2 trajectoires d'abeilles avec et sans puce RFID.	143
4.15	À gauche : photo du système RFID constitué de 2 rangées de 5 lecteurs. À droite : schéma du système illustrant la disposition et la numérotation des lecteurs. Les abeilles possèdent une grande liberté de mouvement entre les 2 rangées de lecteurs.	144
4.16	Signatures de trajectoires pucées VS non-pucées.	145
4.17	Modèles pour la classification des trajectoires d'abeilles et de frelons.	147
4.18	Étude du mouvement sur la planche d'envol relativement au niveau de pression exercée par le frelon asiatique sur la colonie.	148
A.1	Capture d'écran de l'outil d'annotation de cibles.	157
A.2	Capture d'écran de l'outil d'annotation des trajectoires.	158
A.3	Capture d'écran de l'outil de détournage de la planche d'envol.	159
C.1	Exemple de fiche de suivi des acquisitions vidéo journalières.	163
D.1	Ordre de grandeur des spécimens étudiés.	165
D.2	Spécimens vus par la caméra stéréovision TYZX G3 Evo placée à des distances bornées entre 42.5 cm et 60 cm.	166

Liste des tableaux

1	Verrous scientifiques et technologiques de l'étude et aperçu des solutions proposées dans ce manuscrit.	11
1.1	Comparaison des spécifications de 6 caméras TOF et Stéréovision.	20
1.2	Résumé des forces et faiblesses des deux systèmes "TOF CamBoard Nano" et "Stéréovision G3 Evo".	23
1.3	Domaine d'utilisation de la caméra stéréovision TYZX G3 Evo concernant la profondeur Z.	24
1.4	Résultats de la segmentation HIDS par capture (conditions variables).	42
1.5	Synthèse des paramètres (et valeurs par défaut) utilisés dans notre méthode de segmentation HIDS.	44
2.1	Synthèse des problèmes applicatifs identifiés dans cette section 2.4.1 et des solutions proposées.	70
4.1	Principaux facteurs à retenir pour chaque étape de la chaîne de traitement. . .	140

Liste des Algorithmes

1	Mécanisme de reprojexion pour la gestion des observations sans profondeur . .	62
2	Algorithme pour la mise en pratique du clustering par DP.	93
3	Algorithme SMA "Colonie d'abeilles"	162

Bibliographie

- [Adelson 1992] E.H. Adelson et J.Y.A Wang. *Single lens stereo with a plenoptic camera*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 14, no. 2, pages 99–106, Février 1992. (Cit  en page 19.)
- [Alcantarilla 2012] Pablo Fern ndez Alcantarilla, Adrien Bartoli et Andrew J. Davison. *KAZE Features*. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato et Cordelia Schmid,  diteurs, Computer Vision - ECCV 2012, volume 7577 of *Lecture Notes in Computer Science*, pages 214–227. Springer Berlin Heidelberg, 2012. (Cit  en page 26.)
- [Arca 2014] Mariangela Arca, Alexandros Papachristoforou, Florence Mougel, Agn s Rortais, Karine Monceau, Olivier Bonnard, Pascal Tardy, Denis Thi ry, Jean-Fran ois Silvain et G rard Arnold. *Defensive behaviour of Apis mellifera against Vespa velutina in France : Testing whether European honeybees can develop an effective collective defence against a new predator*. Behavioural Processes, vol. 106, no. 0, pages 122 – 129, 2014. (Cit  en page 6.)
- [Arulampalam 2004] B.R.S. Arulampalam. Beyond the kalman filter : Particle filters for tracking applications. Artech House, 2004. (Cit  en pages 51, 53 et 54.)
- [Balch 2001] Tucker Balch, Zia Khan et Manuela Veloso. *Automatically Tracking and Analyzing the Behavior of Live Insect Colonies*. In Proceedings of the Fifth International Conference on Autonomous Agents, AGENTS ’01, pages 521–528, New York, NY, USA, 2001. ACM. (Cit  en page 28.)
- [Banzhaf 1998] Wolfgang Banzhaf, Peter Nordin, Robert E Keller et Frank D Francone. Genetic programming : an introduction, volume 1. Morgan Kaufmann San Francisco, 1998. (Cit  en page 144.)
- [Bar-Shalom 1975] Yaakov Bar-Shalom et Edison Tse. *Tracking in a cluttered environment with probabilistic data association*. Automatica, vol. 11, no. 5, pages 451 – 460, 1975. (Cit  en page 53.)
- [Bashir 2006] FaisalI. Bashir, AshfaqA. Khokhar et Dan Schonfeld. *View-invariant motion trajectory-based activity classification and recognition*. Multimedia Systems, vol. 12, no. 1, pages 45–54, 2006. (Cit  en page 101.)
- [Bay 2006] Herbert Bay, Tinne Tuytelaars et Luc Van Gool. *SURF : Speeded Up Robust Features*. In Ale  Leonardis, Horst Bischof et Axel Pinz,  diteurs, Computer Vision - ECCV 2006, volume 3951 of *Lecture Notes in Computer Science*, pages 404–417. Springer Berlin Heidelberg, 2006. (Cit  en page 26.)
- [Beal 2006] Matthew Beal et Praveen Krishnamurthy. *Gene expression time course clustering with countably infinite hidden Markov models*. In Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence UAI, 2006. (Cit  en page 86.)
- [Bender 2003] Susan FA Bender, Phil J Rodacy, Randal L Schmitt, Philip J Hargis Jr, Mark S Johnson, James R Klarkowski, Glen I Magee et Gary L Bender. *Tracking Honey Bees Using LIDAR (Light Detection and Ranging) Technology*. Rapport technique, Sandia National Laboratories, Janvier 2003. (Cit  en page 123.)

- [Beyan 2013] Cigdem Beyan et Robert Fisher. *Detection of Abnormal Fish Trajectories Using a Clustering Based Hierarchical Classifier*. In Proceedings of the British Machine Vision Conference. BMVA Press, 2013. (Cit  en page 100.)
- [Black 2003] James Black, Tim Ellis et Paul Rosin. *A Novel Method for Video Tracking Performance Evaluation*. In Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS, pages 125–132, 2003. (Cit  en page 75.)
- [Blackman 1986] Samuel Blackman. Multiple-target tracking with radar applications. Artech House, Dedham, MA, 1986. (Cit  en page 55.)
- [Blackman 1999] S.S. Blackman et R. Popoli. Design and analysis of modern tracking systems. Artech House radar library. Artech House, 1999. (Cit  en pages 52, 57 et 58.)
- [Blackwell 1973] David Blackwell et James B. MacQueen. *Ferguson Distributions Via Polya Urn Schemes*. The Annals of Statistics, vol. 1, no. 2, pages 353–355, 03 1973. (Cit  en page 91.)
- [Blois 2011] J. Blois. Vid osurveillance d’abeilles, comptage d’entr es/sorties   l’entr e de la ruche. Master’s thesis, Universit  de La Rochelle, France, 2011. (Cit  en pages 6, 17 et 123.)
- [Blom 1988] H. A P Blom et Y. Bar-Shalom. *The interacting multiple model algorithm for systems with Markovian switching coefficients*. Automatic Control, IEEE Transactions on, vol. 33, no. 8, pages 780–783, Ao t 1988. (Cit  en page 55.)
- [Blum 2012] M. Blum, J.T. Springenberg, J. Wulfin et M. Riedmiller. *A learned feature descriptor for object recognition in RGB-D data*. In Robotics and Automation (ICRA), 2012 IEEE International Conference on, pages 1298–1303, Mai 2012. (Cit  en page 27.)
- [Bo 2013] Liefeng Bo, Xiaofeng Ren et Dieter Fox. *Unsupervised Feature Learning for RGB-D Based Object Recognition*. In Jaydev P. Desai, Gregory Dudek, Oussama Khatib et Vijay Kumar,  diteurs, Experimental Robotics, volume 88 of *Springer Tracts in Advanced Robotics*, pages 387–402. Springer International Publishing, 2013. (Cit  en page 27.)
- [Boers 2003] Y. Boers et J.N. Driessen. *Interacting multiple model particle filter*. Radar, Sonar and Navigation, IEE Proceedings -, vol. 150, no. 5, pages 344–349, Octobre 2003. (Cit  en page 55.)
- [Brillet 2002] Charles Brillet, Gene E. Robinson, R. Bues et Yves Le Conte. *Racial Differences in Division of Labor in Colonies of the Honey Bee (Apis mellifera)*. Ethology, vol. 108, no. 2, pages 115–126, 2002. (Cit  en page 122.)
- [Buchanan 2006] Aeron M. Buchanan et Andrew W. Fitzgibbon. *Interactive Feature Tracking using K-D Trees and Dynamic Programming*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 626–633. IEEE, Janvier 2006. (Cit  en page 120.)
- [Campbell 2008] J. Campbell, L. Mummert et R. Sukthankar. *Video Monitoring of Honey Bee Colonies at the Hive Entrance*. In Proceedings of Workshop on Visual observation and analysis of animal and insect behavior (ICPR), 2008. (Cit  en pages 17, 28 et 123.)
- [Chauvin 1976] R. Chauvin. *Sur la mesure de l’activit  des Abeilles au trou de vol d’une ruche   dix cadres*. Insectes Sociaux, vol. 23, no. 1, pages 75–81, 1976. (Cit  en page 122.)

- [Chen 2012] Chiu Chen, En-Cheng Yang, Joe-Air Jiang et Ta-Te Lin. *An imaging system for monitoring the in-and-out activity of honey bees*. Computers and Electronics in Agriculture, vol. 89, no. 0, pages 100 – 109, 2012. (Cité en pages 27 et 142.)
- [Chiron 2013a] Guillaume Chiron, Petra Gomez-Krämer et Michel Menard. *Detecting and tracking honeybees in 3D at the beehive entrance using stereo vision*. EURASIP Journal on Image and Video Processing, vol. 2013, no. 1, page 59, 2013. (Cité en pages 53, 123 et 152.)
- [Chiron 2013b] Guillaume Chiron, Petra Gomez-Krämer, Michel Ménard et Fabrice Requier. *3D Tracking of Honeybees Enhanced by Environmental Context*. In Alfredo Petrosino, éditeur, Image Analysis and Processing - ICIAP 2013, volume 8156 of *Lecture Notes in Computer Science*, pages 702–711. Springer Berlin Heidelberg, 2013. (Cité en pages 63, 70, 123 et 152.)
- [Chiron 2014a] Guillaume Chiron, Petra Gomez-Krämer et Michel Ménard. *3D tracking based on possibilities rather than probabilities. Application to flying honeybees at the beehive entrance*. In Workshop Visual observation and analysis of Vertebrate And Insect Behavior (VAIB), in conjunction with ICPR, 2014, volume 1, pages 1–1, 2014. (Cité en pages 67, 68, 123 et 152.)
- [Chiron 2014b] Guillaume Chiron, Petra Gomez-Krämer et Michel Ménard. *Approche Bayésienne non-paramétrique pour la découverte de comportements émergents à partir de trajectoires*. In Actes de la conférence RFIA 2014, France, Juin 2014. (Cité en pages 101, 110 et 115.)
- [Chiron 2014c] Guillaume Chiron, Petra Gomez-Krämer et Michel Ménard. *Discovering Emergent Behaviors from Tracks Using Hierarchical Non-parametric Bayesian Methods*. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR), 2014. (Cité en pages 101, 105, 110 et 115.)
- [Cleveland 1981] William S. Cleveland. *LOWESS : A Program for Smoothing Scatterplots by Robust Locally Weighted Regression*. The American Statistician, vol. 35, no. 1, 1981. (Cité en page 72.)
- [Decourtye 2011] Axel Decourtye, James Devillers, Pierrick Aupinel, François Brun, Camille Bagnis, Julie Fourrier et Monique Gauthier. *Honeybee tracking with microchips : a new methodology to measure the effects of pesticides*. Ecotoxicology, vol. 20, no. 2, pages 429–437, 2011. (Cité en page 122.)
- [Dockstader 2001] S.L. Dockstader et AM. Tekalp. *Multiple camera fusion for multi-object tracking*. In Multi-Object Tracking, 2001. Proceedings. 2001 IEEE Workshop on, pages 95–102, 2001. (Cité en page 19.)
- [Dorffer 2014] Clément Dorffer. *Caractérisation de mouvements de masses par analyse du flot optique*. Master’s thesis, Université de La Rochelle, France, 2014. (Cité en page 147.)
- [Doucet 2001] A. Doucet, N. de Freitas et N. Gordon. *Sequential monte carlo methods in practice*. Information Science and Statistics. Springer, 2001. (Cité en pages 51 et 53.)
- [Dovera 2011] Laura Dovera et Ernesto Della Rossa. *Multimodal ensemble Kalman filtering using Gaussian mixture models*. Computational Geosciences, vol. 15, no. 2, pages 307–323, 2011. (Cité en page 53.)
- [Estivill-Castro 2003] V. Estivill-Castro, D. Lattin, F. Suraweera et V. Vithanage. *Tracking bees - a 3D, outdoor small object environment*. In Image Processing, 2003. ICIAP 2003.

- Proceedings. 2003 International Conference on, volume 3, pages III–1021–4 vol.2, Septembre 2003. (Cité en page 124.)
- [Evensen 2003] Geir Evensen. *The ensemble Kalman filter : Theoretical formulation and practical implementation*. Ocean dynamics, vol. 53, no. 4, pages 343–367, 2003. (Cité en page 53.)
- [Fei-Fei 2005] L. Fei-Fei et P. Perona. *A Bayesian hierarchical model for learning natural scene categories*. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 524–531 vol. 2, Juin 2005. (Cité en page 86.)
- [Feldman 2004] Adam Feldman et Tucker Balch. *Representing Honey Bee Behavior for Recognition Using Human Trainable Models*. Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems, vol. 12, no. 3-4, pages 241–250, Septembre 2004. (Cité en pages 87 et 123.)
- [Ferguson 1973] Thomas S. Ferguson. *A Bayesian Analysis of Some Nonparametric Problems*. The Annals of Statistics, vol. 1, no. 2, pages 209–230, 03 1973. (Cité en page 91.)
- [Figueroa 2009] Jr. R.B. Figueroa, A. J. Jacildo et Rabajante J.F. *Simulation of bee foraging behavior using Biroi Preference Algorithm*. In Oral presentation in the 15th BEENET Conference and Techno-fora., pages 57–64, 2009. (Cité en page 110.)
- [Fox 2007] Emily B Fox, Erik B Sudderth et Alan S Willsky. *Hierarchical Dirichlet processes for tracking maneuvering targets*. In Information Fusion, 2007 10th International Conference on, pages 1–8. IEEE, 2007. (Cité en page 86.)
- [Fox 2008] Emily B. Fox, Erik B. Sudderth, Michael I. Jordan et Alan S. Willsky. *An HDP-HMM for Systems with State Persistence*. In Proceedings of the 25th International Conference on Machine Learning, ICML '08, pages 312–319, New York, NY, USA, 2008. ACM. (Cité en page 95.)
- [Gemeiner 2009] P. Gemeiner, P. Jojic et M. Vincze. *Selecting good corners for structure and motion recovery using a time-of-flight camera*. In Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on, pages 5711–5716, Octobre 2009. (Cité en page 27.)
- [Goktuk 2005] S.B. Goktuk et A Rafii. *An Occupant Classification System Eigen Shapes or Knowledge-Based Features*. In Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, pages 57–57, Juin 2005. (Cité en page 27.)
- [Goldstein 1973] G.B. Goldstein. *False-Alarm Regulation in Log-Normal and Weibull Clutter*. Aerospace and Electronic Systems, IEEE Transactions on, vol. AES-9, no. 1, pages 84–92, Janvier 1973. (Cité en page 142.)
- [Gordon 1993] N.J. Gordon, D.J. Salmond et A F M Smith. *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*. Radar and Signal Processing, IEE Proceedings F, vol. 140, no. 2, pages 107–113, Avril 1993. (Cité en pages 49 et 51.)
- [Gorelick 2007] Lena Gorelick, Moshe Blank, Eli Shechtman, Michal Irani et Ronen Basri. *Actions As Space-Time Shapes*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 12, pages 2247–2253, Décembre 2007. (Cité en page 87.)

- [Griffin 2006] J. E Griffin et M. F. J Steel. *Order-Based Dependent Dirichlet Processes*. Journal of the American Statistical Association, vol. 101, no. 473, pages 179–194, 2006. (Cit  en page 86.)
- [Harris 1988] C. Harris et M. Stephens. *A Combined Corner and Edge Detector*. In Proceedings of the 4th Alvey Vision Conference, pages 147–151, 1988. (Cit  en page 26.)
- [Hendriks 2012] Cris L. Hendriks, Ziquan Yu, Antoine Lecocq, Teatske Bakker, Barbara Locke et Olle Terenius. *Identifying all individuals in a honeybee hive : progress towards mapping all social interactions*. In Visual observation and analysis of animal and insect behavior, pages 5–8, 2012. (Cit  en pages 27 et 142.)
- [Henry 2012] Micka l Henry, Maxime B guin, Fabrice Requier, Orianne Rollin, Jean-Fran ois Odoux, Pierrick Aupinel, Jean Aptel, Sylvie Tchamitchian et Axel Decourtye. *A Common Pesticide Decreases Foraging Success and Survival in Honey Bees*. Science, vol. 336, no. 6079, pages 348–350, 2012. (Cit  en pages 5 et 6.)
- [Hongeng 2001] Somboon Hongeng et R. Nevatia. *Multi-agent event recognition*. In Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, volume 2, pages 84–91 vol.2, 2001. (Cit  en page 87.)
- [Hospedales 2009] Timothy Hospedales, Shaogang Gong et Tao Xiang. *A Markov Clustering Topic Model for mining behaviour in video*. In Computer Vision, 2009 IEEE 12th International Conference on, pages 1165–1172, Septembre 2009. (Cit  en pages 88 et 97.)
- [Howard 2007] A Howard, LH Matthies, A Huertas, M Bajracharya et A Rankin. *Detecting pedestrians with stereo vision : safe operation of autonomous ground vehicles in dynamic environments*. In Proceedings of the 13th Int. Symp. of Robotics Research, 2007. (Cit  en page 27.)
- [Huang 2009] Shi-Guo Huang, Xiao-lin Li, Ming-quan Zhou et Guo-hua Geng. *SURF-Based Multi-scale Resolution Histogram for Insect Recognition*. In Artificial Intelligence and Computational Intelligence, 2009. AICI '09. International Conference on, volume 2, pages 445–448, Novembre 2009. (Cit  en page 27.)
- [Intille 1999] Stephen S. Intille et Aaron F. Bobick. *A Framework for Recognizing Multi-agent Action from Visual Evidence*. In Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence, AAAI '99/IAAI '99, pages 518–525, Menlo Park, CA, USA, 1999. American Association for Artificial Intelligence. (Cit  en page 87.)
- [Jaward 2006] M. Jaward, L. Mihaylova, N. Canagarajah et D. Bull. *Multiple object tracking using particle filters*. In Aerospace Conference, 2006 IEEE, pages 8 pp.–, 2006. (Cit  en page 55.)
- [Jiwon 2009] Shin Jiwon, Stefan Gachter, A Harati, C. Pradalier et R. Siegwart. *Object classification based on a geometric grammar with a range camera*. In Robotics and Automation, 2009. ICRA '09. IEEE International Conference on, pages 2443–2448, Mai 2009. (Cit  en page 27.)
- [Julier 1997] Simon J. Julier et Jeffrey K. Uhlmann. *New extension of the Kalman filter to nonlinear systems*. In Proceedings of Signal Processing, Sensor Fusion, and Target Recognition VI (SPIE), volume 3068, pages 182–193, 1997. (Cit  en page 52.)

- [Kalman 1960] R. E. Kalman. *A New Approach to Linear Filtering and Prediction Problems*. Transactions of the ASME - Journal of Basic Engineering, no. 82 (Series D), pages 35–45, 1960. (Cité en pages 49 et 52.)
- [Kameda 1996] Y. Kameda et M. Minoh. *A human motion estimation method using 3-successive video frames*. In Proceedings of International Conference on Virtual Systems, Janvier 1996. (Cité en page 26.)
- [Khan 2004a] Zia Khan, T. Balch et F. Dellaert. *A Rao-Blackwellized particle filter for EigenTracking*. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, pages II–980–II–986 Vol.2, Juin 2004. (Cité en pages 27 et 123.)
- [Khan 2004b] Zia Khan, Tucker Balch et Frank Dellaert. *An MCMC-Based Particle Filter for Tracking Multiple Interacting Targets*. In Tomás Pajdla et Jirí Matas, éditeurs, Computer Vision - ECCV 2004, volume 3024 of *Lecture Notes in Computer Science*, pages 279–290. Springer Berlin Heidelberg, 2004. (Cité en pages 54 et 55.)
- [Khan 2005] Zia Khan, T. Balch et F. Dellaert. *MCMC-based particle filtering for tracking a variable number of interacting targets*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 11, pages 1805–1819, Novembre 2005. (Cité en page 87.)
- [Kimura 2011] Toshifumi Kimura, Mizue Ohashi, Ryuichi Okada et Hidetoshi Ikeno. *A new approach for the simultaneous tracking of multiple honeybees for analysis of hive behavior*. Apidologie, vol. 42, no. 5, pages 607–617, 2011. (Cité en pages 28 et 123.)
- [Kimura 2014] Toshifumi Kimura, Mizue Ohashi, Karl Crailsheim, Thomas Schmickl, Ryuichi Okada, Gerald Radspieler et Hidetoshi Ikeno. *Development of a New Method to Track Multiple Honey Bees with Complex Behaviors on a Flat Laboratory Arena*. PLoS ONE, vol. 9, no. 1, page e84656, 01 2014. (Cité en pages 28 et 123.)
- [Kong 1994] Augustine Kong, Jun S. Liu et Wing H. Wong. *Sequential Imputations and Bayesian Missing Data Problems*. Journal of the American Statistical Association, vol. 89, no. 425, pages 278–288, Mars 1994. (Cité en page 51.)
- [Kuettel 2010] D. Kuettel, M.D. Breitenstein, L. Van Gool et V. Ferrari. *What's going on? Discovering spatio-temporal dependencies in dynamic scenes*. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 1951–1958, Juin 2010. (Cité en pages 88 et 97.)
- [Kuhn 1955] Harold W. Kuhn. *The Hungarian method for the assignment problem*. Naval research logistics quarterly, vol. 2, no. 1-2, pages 83–97, 1955. (Cité en page 57.)
- [Lemmens 2008] Nyree Lemmens, Steven de Jong, Karl Tuyls et Ann Nowé. *Bee Behaviour in Multi-agent Systems*. In Karl Tuyls, Ann Nowe, Zahia Guessoum et Daniel Kudenko, éditeurs, Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning, volume 4865 of *Lecture Notes in Computer Science*, pages 145–156. Springer Berlin Heidelberg, 2008. (Cité en page 110.)
- [Li 2009] Xiao lin Li, Shi-Guo Huang, Ming-quan Zhou et Guo-hua Geng. *KNN-Spectral Regression LDA for Insect Recognition*. In Information Science and Engineering (ICISE), 2009 1st International Conference on, pages 1315–1318, Décembre 2009. (Cité en page 27.)

- [Liu 2001] Jane Liu et Mike West. *Combined Parameter and State Estimation in Simulation-Based Filtering*. In Arnaud Doucet, Nando de Freitas et Neil Gordon, éditeurs, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, pages 197–223. Springer New York, 2001. (Cité en page 53.)
- [Liu 2012] Ye Liu, Hui Li et YanQiu Chen. *Automatic Tracking of a Large Number of Moving Targets in 3D*. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato et Cordelia Schmid, éditeurs, *Computer Vision - ECCV 2012*, volume 7575 of *Lecture Notes in Computer Science*, pages 730–742. Springer Berlin Heidelberg, 2012. (Cité en page 28.)
- [Lowe 1999] D.G. Lowe. *Object recognition from local scale-invariant features*. In *Computer Vision*, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2, pages 1150–1157 vol.2, 1999. (Cité en page 26.)
- [MacEachern 1999] S. MacEachern. *Dependent Nonparametric Processes*. In *ASA Proceedings of the Section on Bayesian Statistical Science*, 1999. (Cité en page 95.)
- [Mahler 2003] R.P.S. Mahler. *Multitarget Bayes filtering via first-order multitarget moments*. *Aerospace and Electronic Systems*, IEEE Transactions on, vol. 39, no. 4, pages 1152–1178, Octobre 2003. (Cité en page 49.)
- [Maitra 2009] P. Maitra, S. Schneider et M.C. Shin. *Robust bee tracking with adaptive appearance template and geometry-constrained resampling*. In *Applications of Computer Vision (WACV)*, 2009 Workshop on, pages 1–6, Décembre 2009. (Cité en pages 27 et 123.)
- [McFarlane 1995] N.J.B. McFarlane et C.P. Schofield. *Segmentation and tracking of piglets in images*. *Machine Vision and Applications*, vol. 8, no. 3, pages 187–193, 1995. (Cité en page 26.)
- [Meyer 1994] Fernand Meyer. *Topographic distance and watershed lines*. *Signal Processing*, vol. 38, no. 1, pages 113 – 125, 1994. *Mathematical Morphology and its Applications to Signal Processing*. (Cité en page 43.)
- [Miller 1997] M.L. Miller, H.S. Stone et Ingemar J. Cox. *Optimizing Murty’s ranked assignment method*. *Aerospace and Electronic Systems*, IEEE Transactions on, vol. 33, no. 3, pages 851–862, Juillet 1997. (Cité en page 58.)
- [Miranda 2012] B. Miranda, J. Salas et P. Vera. *Bumblebees Detection and Tracking*. In *Proceedings of Workshop on Visual observation and analysis of animal and insect behavior (ICPR)*, 2012. (Cité en page 27.)
- [Morris 2008] B.T. Morris et M.M. Trivedi. *A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance*. *Circuits and Systems for Video Technology*, IEEE Transactions on, vol. 18, no. 8, pages 1114–1127, Août 2008. (Cité en page 87.)
- [Niebles 2008] JuanCarlos Niebles, Hongcheng Wang et Li Fei-Fei. *Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words*. *International Journal of Computer Vision*, vol. 79, no. 3, pages 299–318, 2008. (Cité en page 86.)
- [Nummiaro 2003] Katja Nummiaro, Esther Koller-Meier, TomÁš Svoboda, Daniel Roth et Luc Van Gool. *Color-Based Object Tracking in Multi-camera Environments*. In Bernd Michaelis et Gerald Krell, éditeurs, *Pattern Recognition*, volume 2781 of *Lecture Notes in Computer Science*, pages 591–599. Springer Berlin Heidelberg, 2003. (Cité en page 19.)

- [Oh 2008] Sang Min Oh, James M.M. Rehg, Tucker Balch et Frank Dellaert. *Learning and Inferring Motion Patterns using Parametric Segmental Switching Linear Dynamic Systems*. International Journal of Computer Vision, vol. 77, no. 1-3, pages 103–124, 2008. (Cité en page 55.)
- [Oliver 2000] N.M. Oliver, B. Rosario et AP. Pentland. *A Bayesian computer vision system for modeling human interactions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 8, pages 831–843, Août 2000. (Cité en page 87.)
- [Orbanz 2011] P. Orbanz et Y. W. Teh. *Modern Bayesian nonparametrics*. Presented In Neural Information Processing Systems, 2011. (Cité en page 88.)
- [Osborne 1999] J.L. Osborne, S.J. Clark, R.J. Morris, I.H. Williams, J.R. Riley, A.D. Smith, D.R. Reynolds et A.S. Edwards. *A landscape-scale study of bumble bee foraging range and constancy, using harmonic radar*. Journal of Applied Ecology, vol. 36, no. 4, pages 519–533, 1999. (Cité en page 6.)
- [Oudjane 1999] Nadia Oudjane et Christian Musso. *Multiple model particle filter*. In 17 Colloque sur le traitement du signal et des images, FRA, 1999. GRETSI, Groupe d’Etudes du Traitement du Signal et des Images, 1999. (Cité en page 55.)
- [Piatti 2011] Dario Piatti. *Time-of-Flight cameras : test, calibration and multi-frame registration for automatic 3D object reconstruction*. PhD thesis, Politecnico di Torino, Italy, 2011. (Cité en page 22.)
- [Pitman 2002] Jim Pitman. *Poisson-Dirichlet and GEM Invariant Distributions for Split-and-Merge Transformations of an Interval Partition*. Combinatorics, Probability and Computing, vol. 11, pages 501–514, 9 2002. (Cité en page 91.)
- [Pitt 1999] Michael K. Pitt et Neil Shephard. *Filtering via simulation : Auxiliary particle filters*. Journal of the American statistical association, vol. 94, no. 446, pages 590–599, 1999. (Cité en page 53.)
- [Portelli 2011] Geoffrey Portelli, Franck Ruffier, Frédéric L. Roubieu et Nicolas Franceschini. *Honeybees’ Speed Depends on Dorsal as Well as Lateral, Ventral and Frontal Optic Flows*. PLoS ONE, vol. 6, no. 5, page e19486, 05 2011. (Cité en page 70.)
- [Raharimanana 2014] Heriniaina Raharimanana. *Modélisation du comportement d’une ruche à partir de données rfid bruitées d’entrées/sorties d’abeilles*. Master’s thesis, Université de La Rochelle, France, 2014. (Cité en page 144.)
- [Rao 2002] Cen Rao, Alper Yilmaz et Mubarak Shah. *View-Invariant Representation and Recognition of Actions*. International Journal of Computer Vision, vol. 50, no. 2, pages 203–226, 2002. (Cité en page 87.)
- [Reda 2013] Khairi Reda, Victor Mateevitsi et Catherine Offord. *A human-computer collaborative workflow for the acquisition and analysis of terrestrial insect movement in behavioral field studies*. EURASIP Journal on Image and Video Processing, vol. 2013, no. 1, 2013. (Cité en page 28.)
- [Reid 1979] D.B. Reid. *An algorithm for tracking multiple targets*. Automatic Control, IEEE Transactions on, vol. 24, no. 6, pages 843–854, Décembre 1979. (Cité en pages 54 et 58.)
- [Requier 2014] Fabrice Requier. *L’écologie de l’abeille domestique en système de céréaliculture intensive*. PhD thesis, Centre d’Études Biologiques de Chizé (CEBC), et Institut national de la recherche agronomique (INRA Poitou-Charentes), 2014. (Cité en page 141.)

- [Riley 1996] JR Riley, AD Smith, DR Reynolds, AS Edwards, JL Osborne, IH Williams, NL Carreck et GM Poppy. *Tracking bees with harmonic radar*. Nature, vol. 379, no. 6560, pages 29–30, 1996. (Cité en page 124.)
- [Rodríguez 2008] Abel Rodríguez, David B Dunson et Alan E Gelfand. *The Nested Dirichlet Process*. Journal of the American Statistical Association, vol. 103, no. 483, pages 1131–1154, 2008. (Cité en page 95.)
- [Schaefer 2012] Andreas T Schaefer et Adam Claridge-Chang. *The surveillance state of behavioral automation*. Current opinion in neurobiology, vol. 22, no. 1, pages 170–176, 2012. (Cité en page 86.)
- [Scharstein 2003] D. Scharstein et R. Szeliski. *High-accuracy stereo depth maps using structured light*. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pages I–195–I–202 vol.1, Juin 2003. (Cité en page 19.)
- [Sethuraman 1994] Jayaram Sethuraman. *A constructive definition of Dirichlet priors*. Statistica Sinica, vol. 4, pages 639–650, 1994. (Cité en page 91.)
- [Simon 2002] D. Simon et Li Chia Tien. *Kalman filtering with state equality constraints*. Aerospace and Electronic Systems, IEEE Transactions on, vol. 38, no. 1, pages 128–136, Janvier 2002. (Cité en page 54.)
- [Sivic 2005] J. Sivic, B.C. Russell, AA Efros, A Zisserman et W.T. Freeman. *Discovering objects and their location in images*. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, volume 1, pages 370–377 Vol. 1, Octobre 2005. (Cité en page 26.)
- [Smith 2007] Keston W. Smith. *Cluster ensemble Kalman filter*. Tellus A, vol. 59, no. 5, pages 749–757, 2007. (Cité en page 53.)
- [Sobral 2014] Andrews Sobral et Antoine Vacavant. *A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos*. Computer Vision and Image Understanding, vol. 122, no. 0, pages 4 – 21, 2014. (Cité en pages 16 et 26.)
- [Spangler 1969] Hayward G. Spangler. *Photoelectrical Counting of Outgoing and Incoming Honey Bees*. Journal of Economic Entomology, vol. 62, no. 5, pages 1183–1184, 1969. (Cité en page 123.)
- [Spitzen 2013] Jeroen Spitzen, Cornelis W. Spoor, Fabrizio Grieco, Cajo ter Braak, Jacob Beeuwkes, Sjaak P. van Brugge, Sander Kranenbarg, Lucas P. J. J. Noldus, Johan L. van Leeuwen et Willem Takken. *A 3D Analysis of Flight Behavior of Anopheles gambiae sensu stricto Malaria Mosquitoes in Response to Human Odor and Heat*. PLoS ONE, vol. 8, no. 5, page e62995, 05 2013. (Cité en page 6.)
- [Stauffer 2000] Chris Stauffer et W.E.L. Grimson. *Learning patterns of activity using real-time tracking*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 8, pages 747–757, Août 2000. (Cité en page 87.)
- [Steiger 2009] Olivier Steiger, Stephan Weiss et Judith Felder. *Real-time understanding of 3D video on an embedded system*. Proc. of the EUSIPCO 2009, 2009. (Cité en pages 27 et 28.)
- [Straw 2010] Andrew D. Straw, Kristin Branson, Titus R. Neumann et Michael H. Dickinson. *Multi-camera real-time three-dimensional tracking of multiple flying animals*. Journal of The Royal Society Interface, vol. 7, no. 49, 2010. (Cité en pages 19 et 28.)

- [Streit 2003] Sebastian Streit, Fiola Bock, Christian W.W. Pirk et J  rgen Tautz. *Automatic life-long monitoring of individual insect behaviour now possible*. *Zoology*, vol. 106, no. 3, pages 169 – 171, 2003. (Cit   en pages 6, 122 et 145.)
- [Struye 1994] M. H. Struye, H. J. Mortier, G. Arnold, C. Miniggio et R. Borneck. *Microprocessor-controlled monitoring of honeybee flight activity at the hive entrance*. *Apidologie*, vol. 25, no. 4, pages 384–395, 1994. (Cit   en page 123.)
- [Sudderth 2005] Erik B. Sudderth, Antonio Torralba, William T. Freeman et Alan S. Willsky. *Learning Hierarchical Models of Scenes, Objects, and Parts*. In *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2, ICCV '05*, pages 1331–1338, Washington, DC, USA, 2005. IEEE Computer Society. (Cit   en page 86.)
- [Sumpter 2004] D. J. T. Sumpter et S. J. Martin. *The dynamics of virus epidemics in Varroa-infested honey bee colonies*. *Journal of Animal Ecology*, vol. 73, no. 1, pages 51–63, 2004. (Cit   en page 6.)
- [Swadzba 2008] A Swadzba, N. Beuter, J. Schmidt et G. Sagerer. *Tracking objects in 6D for reconstructing static scenes*. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–7, Juin 2008. (Cit   en page 27.)
- [Teh 2006] Yee Whye Teh, Michael I Jordan, Matthew J Beal et David M Blei. *Hierarchical dirichlet processes*. *Journal of the american statistical association*, vol. 101, no. 476, 2006. (Cit   en pages 88, 91 et 94.)
- [Teh 2007] Yee Whye Teh. *Dirichlet processes : Tutorial and practical course*. 2007. (Cit   en page 88.)
- [Teh 2010] Yee Whye Teh et Michael I Jordan. *Hierarchical Bayesian nonparametric models with applications*. *Bayesian nonparametrics*, pages 158–207, 2010. (Cit   en page 88.)
- [Theriault 2010] D. Theriault, Z. Wu, N. Hristov, S. Swartz, K. Breuer, T. Kunz et M. Betke. *Reconstruction and analysis of 3D trajectories of Brazilian free-tailed bats in flight*. Rapport technique, CS Department, Boston University, 2010. (Cit   en pages 19 et 28.)
- [Tukey 1962] John W. Tukey. *The Future of Data Analysis*. *The Annals of Mathematical Statistics*, vol. 33, no. 1, pages 1–67, 03 1962. (Cit   en page 72.)
- [Tweed 2002] David Tweed et Andrew Calway. *Tracking Many Objects Using Subordinated Condensation*. In *Proceedings of the British Machine Vision Conference*, pages 283–292, 2002. (Cit   en page 55.)
- [Veeraraghavan 2008] Ashok Veeraraghavan, Rama Chellappa et Mandyam Srinivasan. *Shape-and-Behavior Encoded Tracking of Bee Dances*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 3, pages 463–476, Mars 2008. (Cit   en pages 6, 27, 55 et 123.)
- [Vidau 2011] Cyril Vidau, Marie Diogon, Julie Aufauvre, R  gis Fontbonne, Bernard Vigu  s, Jean-Luc Brunet, Catherine Texier, David G. Biron, Nicolas Blot, Hicham El Alaoui, Luc P. Belzunces et Fr  d  ric Delbac. *Exposure to Sublethal Doses of Fipronil and Thiacloprid Highly Increases Mortality of Honeybees Previously Infected by Nosema ceranae*. *PLoS ONE*, vol. 6, no. 6, page e21550, 06 2011. (Cit   en page 6.)

- [Viola 2001] P. Viola et M. Jones. *Rapid object detection using a boosted cascade of simple features*. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I-511–I-518 vol.1, 2001. (Cité en pages 26 et 43.)
- [Von Frisch 1964] K. Von Frisch. Bees : their vision, chemical senses, and language. Cornell Univ. Press, 1964. (Cité en page 20.)
- [Von Frisch 1965] Karl Von Frisch. Tanzsprache und orientierung der bienen. Berlin, Heidelberg : Springer, 1965. (Cité en page 6.)
- [Walker 1999] Stephen G. Walker, Paul Damien, PuruShottam W. Laud et Adrian F. M. Smith. *Bayesian Nonparametric Inference for Random Distributions and Related Functions*. Journal of the Royal Statistical Society : Series B (Statistical Methodology), vol. 61, no. 3, pages 485–527, 1999. (Cité en page 88.)
- [Wang 2006] Xiaogang Wang, Kinh Tieu et Eric Grimson. *Learning Semantic Scene Models by Trajectory Analysis*. In Ales Leonardis, Horst Bischof et Axel Pinz, éditeurs, Computer Vision - ECCV 2006, volume 3953 of *Lecture Notes in Computer Science*, pages 110–123. Springer Berlin Heidelberg, 2006. (Cité en page 87.)
- [Wang 2009] Xiaogang Wang, Xiaoxu Ma et W.E.L. Grimson. *Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 3, pages 539–555, Mars 2009. (Cité en pages 88 et 97.)
- [Wang 2011] Xiaogang Wang, W Eric L Grimson et Carl-Fredrik Westin. *Tractography segmentation using a hierarchical Dirichlet processes mixture model*. NeuroImage, vol. 54, no. 1, pages 290–302, 2011. (Cité en page 88.)
- [Wang 2013] Xiaogang Wang. *Intelligent multi-camera video surveillance : A review*. Pattern Recognition Letters, vol. 34, no. 1, pages 3 – 19, 2013. Extracting Semantics from Multi-Spectrum Video. (Cité en page 19.)
- [Wilensky 1997] U Wilensky. *NetLogo ants model*, 1997. (Cité en page 105.)
- [Williamson 2013] Sally M Williamson, Christopher Moffat, Martha Gomersall, Nastja Saranzewa, Christopher Connolly et Geraldine A Wright. *Exposure to acetylcholinesterase inhibitors alters the physiology and motor function of honeybees*. Frontiers in Physiology, vol. 4, no. 13, 2013. (Cité en page 123.)
- [Wren 1997] C.R. Wren, A Azarbajejani, T. Darrell et AP. Pentland. *Pfinder : real-time tracking of the human body*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, no. 7, pages 780–785, Juillet 1997. (Cité en page 26.)
- [Wulsin 2012] Drausin Wulsin, Brian Litt et Shane T. Jensen. *A Hierarchical Dirichlet Process Model with Multiple Levels of Clustering for Human EEG Seizure Modeling*. In John Langford et Joelle Pineau, éditeurs, Proceedings of the 29th International Conference on Machine Learning (ICML-12), pages 57–64, New York, NY, USA, 2012. ACM. (Cité en pages 11, 88, 95, 96, 97 et 98.)
- [Yang 2011] Jianyu Yang, YF Li, Keyi Wang, Yuan Wu, Giuseppe Altieri et Massimo Scalia. *Mixed signature : an invariant descriptor for 3D motion trajectory perception and recognition*. Mathematical Problems in Engineering, vol. 2012, 2011. (Cité en page 100.)

- [Yilmaz 2006] Alper Yilmaz, Omar Javed et Mubarak Shah. *Object Tracking : A Survey*. ACM Computing Surveys, vol. 38, no. 4, Décembre 2006. (Cité en pages 47, 48, 56 et 173.)
- [Yin 2007] Fei Yin, Dimitrios Makris et Sergio A Velastin. *Performance evaluation of object tracking algorithms*. In 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2007), Rio de Janeiro, Brazil, 2007. (Cité en page 75.)
- [Yuefang 2007] Gao Yuefang, Song Hongzhi, Tian Xuhong et Chen Yan. *Identification Algorithm of Winged Insects Based on Hybrid Moment Invariants*. In Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on, pages 531–534, Juillet 2007. (Cité en page 27.)
- [Zelnik-Manor 2001] L. Zelnik-Manor et M. Irani. *Event-based analysis of video*. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 2, pages II–123–II–130 vol.2, 2001. (Cité en page 87.)
- [Zhong 2004] Hua Zhong, Jianbo Shi et M. Visontai. *Detecting unusual activity in video*. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, pages II–819–II–826 Vol.2, Juin 2004. (Cité en page 87.)
- [Zhou 2006] Quming Zhou et JK Aggarwal. *Object tracking in an outdoor environment using fusion of features and cameras*. Image and Vision Computing, vol. 24, no. 11, pages 1244–1255, 2006. (Cité en page 19.)
- [Zhu 2011] Xiaobin Zhu, Jing Liu, Jinqiao Wang, Wei Fu, Hanqing Lu et Yikai Fang. *Global Trajectory Construction across Multi-cameras via Graph Matching*. In Image and Graphics (ICIG), 2011 Sixth International Conference on, pages 801–806, Août 2011. (Cité en page 19.)
- [Zivkovic 2006] Zoran Zivkovic et Ferdinand van der Heijden. *Efficient adaptive density estimation per image pixel for the task of background subtraction*. Pattern Recognition Letters, vol. 27, no. 7, pages 773 – 780, 2006. (Cité en page 26.)

Système complet d'acquisition vidéo, de suivi de trajectoires et de modélisation comportementale pour des environnements 3D naturellement encombrés. Application à la surveillance apicole.

Ce manuscrit propose une approche méthodologique pour la constitution d'une chaîne complète de vidéosurveillance pour des environnements naturellement encombrés. Nous identifions et levons un certain nombre de verrous méthodologiques et technologiques inhérents : 1) à l'acquisition de séquences vidéo en milieu naturel, 2) au traitement d'images, 3) au suivi multi-cibles, 4) à la découverte et la modélisation de motifs comportementaux récurrents, et 5) à la fusion de données. Le contexte applicatif de nos travaux est la surveillance apicole, et en particulier, l'étude des trajectoires des abeilles en vol devant la ruche. De ce fait, cette thèse se présente également comme une étude de faisabilité et de prototypage dans le cadre des deux projets interdisciplinaires EPERAS et RISQAPI (projets menés en collaboration avec l'INRA Magneraud et le Muséum National d'Histoire Naturelle). Il s'agit pour nous informaticiens et pour les biologistes qui nous ont accompagnés, d'un domaine d'investigation totalement nouveau, pour lequel les connaissances métiers, généralement essentielles à ce genre d'applications, restent encore à définir.

Contrairement aux approches existantes de suivi d'insectes, nous proposons de nous attaquer au problème dans l'espace à trois dimensions grâce à l'utilisation d'une caméra stéréovision haute fréquence. Dans ce contexte, nous détaillons notre nouvelle méthode de détection de cibles appelée segmentation HIDS. Concernant le calcul des trajectoires, nous explorons plusieurs approches de suivi de cibles, s'appuyant sur plus ou moins *a priori*, susceptibles de supporter les conditions extrêmes de l'application (e.g. cibles nombreuses, de petite taille, présentant un mouvement chaotique). Une fois les trajectoires collectées, nous les organisons selon une structure de données hiérarchique et mettons en œuvre une approche Bayésienne non-paramétrique pour la découverte de comportements émergents au sein de la colonie d'insectes. L'analyse exploratoire des trajectoires issues de la scène encombrée s'effectue par classification non supervisée, simultanément sur des niveaux sémantiques différents, et où le nombre de clusters pour chaque niveau n'est pas défini *a priori* mais est estimé à partir des données. Cette approche est dans un premier temps validée à l'aide d'une pseudo-vérité terrain générée par un Système Multi-Agents, puis dans un deuxième temps appliquée sur des données réelles.

Mots clés : stéréovision, segmentation RGB-D, suivi multicibles, modélisation comportementale, approche Bayésienne non-paramétrique, processus hiérarchique de Dirichlet, surveillance apicole, colonie d'abeilles.

Full process of acquisition, multi-target tracking, behavioral modeling for naturally crowded environments. Application to beehives monitoring.

This manuscript provides the basis for a complete chain of videosurveillance for naturally cluttered environments. In the latter, we identify and solve the wide spectrum of methodological and technological barriers inherent to : 1) the acquisition of video sequences in natural conditions, 2) the image processing problems, 3) the multi-target tracking ambiguities, 4) the discovery and the modeling of recurring behavioral patterns, and 5) the data fusion. The application context of our work is the monitoring of honeybees, and in particular the study of the trajectories bees in flight in front of their hive. In fact, this thesis is part a feasibility and prototyping study carried by the two interdisciplinary projects EPERAS and RISQAPI (projects undertaken in collaboration with INRA institute and the French National Museum of Natural History). It is for us, computer scientists, and for biologists who accompanied us, a completely new area of investigation for which the scientific knowledge, usually essential for such applications, are still in their infancy.

Unlike existing approaches for monitoring insects, we propose to tackle the problem in the three-dimensional space through the use of a high frequency stereo camera. In this context, we detail our new target detection method which we called HIDS segmentation. Concerning the computation of trajectories, we explored several tracking approaches, relying on more or less *a priori*, which are able to deal with the extreme conditions of the application (e.g. many targets, small in size, following chaotic movements). Once the trajectories are collected, we organize them according to a given hierarchical data structure and apply a Bayesian nonparametric approach for discovering emergent behaviors within the colony of insects. The exploratory analysis of the trajectories generated by the crowded scene is performed following an unsupervised classification method simultaneously over different levels of semantic, and where the number of clusters for each level is not defined *a priori*, but rather estimated from the data only. This approach is has been validated thanks to a ground truth generated by a Multi-Agent System. Then we tested it in the context of real data.

Keywords : stereovision, RGB-D segmentation, multi-target tracking, behavioral modeling, Bayesian nonparametric approach, hierarchical Dirichlet process, beehive monitoring, honeybee colony.